

Is there a phoenix in the ashes?

A forthcoming report on teaching and research in the Earth sciences is a good prescription for the unavoidable changes that still lie ahead for British universities.

If it can be assumed that the British university system has now passed through the nadir of its fortunes so that, from now on, things cannot get yet worse, and may even improve a little with the government's modest recognition of the need for more higher education, one important recipe for improvement must surely be found in a recommendation for the reorganization of the Earth sciences in Britain, due to be circulated next week by the University Grants Committee. There is every chance that the recipe now offered in this one field may be applicable — and be applied — in others.

The document is the work of a committee under Professor E.R. Oxburgh established only last October. Its job has been to suggest a way of making better use of present academic provision from teaching and research in the Earth sciences, now distributed more or less by chance among 36 departments at various university institutions. Briefly, the committee recommends that there should in future be three kinds of university centres with an interest in the field — some (about ten) equipped both for research and teaching, some equipped for teaching only, but able to provide a full range of instruction, and others that would be able to provide undergraduates, especially early in their academic careers, with understanding of the Earth sciences, but which would not be free-standing academic enterprises. Of necessity, the rounded centres would be fairly large university departments, with perhaps 30 academics. Teaching departments would be roughly half the size, but would be able to aspire to "excellence in teaching". The smallest centres would often be embedded in other departments, such as physics.

Essence

The essence of the Oxburgh committee's argument is that a distribution of academic strength along these lines is not merely a way of using scarce resources economically, but that it is also "good" for the Earth sciences themselves. Wisely, this conclusion hangs not on the usual hand-waving that academic success requires a "critical mass" of academic strength, but on a careful comparison of the numbers of research students in British and US university departments in the field. (Surprisingly, the distributions are not very different; US Earth science departments with mere handfuls of students are commonplace, but Britain is deficient in large departments.) Moreover, it does not condemn small departments out of hand, but rather speaks of their virtues, their informality in particular. Where large departments differ from the small is in their capacity to sustain and service complicated research programmes, their capacity to undertake initiatives from their own resources and, just as important, their provision of a climate in which research students learn as much from each other as from those who are formally paid to teach them. So within an unchanged provision for the Earth sciences in British universities, Oxburgh wants to see money and people taken from most so as to be given to the chosen ten (or even twelve). The argument is persuasive enough that the conclusion will stick even with those who are to be deprived.

If this is the natural pattern, why has it not arisen naturally, in the regular ebb and flow of academic fortunes? Part of the explanation is the influence of the Natural Environmental

Research Council (NERC) on the pattern of research; the traditional practice (followed by other research councils) of awarding individual departments the right to award stipends to research students of their choice has a disproportionate effect on the distribution of larger project funds. The Oxburgh report asks that NERC should keep its practice "under review", but it would be better if NERC (and the other British research councils) were to abandon the practice altogether; why not make awards to students, and let them go where the prospects are most exciting?

Central facilities

Oxburgh also points out that NERC's more recent practice of providing central facilities (mass spectrometry services, for example) at which hard-pressed academics have access to equipment their own institutions could not afford is inconsistent with the pattern of concentration. But research vessels are a different matter while, the argument imaginatively continues, there is a case for planning for some fields of research new to Britain (high-pressure studies, for example) on a European basis. It will be best if NERC, which is sensitive to criticism, regards these criticisms not as a complaint but as a reminder to all the research councils that their support of university research needs to be reassessed, not that their pensioners are so quickly changing.

With caveats such as these, the chances are that British universities will accept that the Oxburgh recipe is both sensible for the Earth sciences and a model for the concentration of resources in other fields. The trouble will come only when particular departments are to be singled out for a particular future. Here the recipe is that those which set their caps on becoming (or remaining) large and comprehensive should formally apply to the grants committee for the resources they will need. The suggested deadline for applications is next January, which will seem like tomorrow to many universities. But that is the pace necessary if worthwhile enterprises are to be rescued from the calamities of the past two decades. Let us hope that those concerned have strong nerves.

Even with good luck, however, it would be a mistake to expect that the simple application of this recipe to other fields would bring about the concentration of resources on centres of excellence that is now so plainly necessary in the British system. For one thing, there is a danger that, on the principle of fair play, universities robbed (as they see it) of a centre in one field should be favoured in another by way of compensation, which would ensure that every university is lopsided. For another, there is a danger (which Oxburgh recognizes) that the paymasters, having created a new pattern, will fail to see the need that smaller university departments should be enabled to grow if they have the talent to do so. Then, whether they like it or not, universities will have to change the pattern of their undergraduate teaching so as to assist the transfer of students from one institution to another. Finally, the artificial gulf between the so-called public and private sectors of higher education (both supported by the central government), which is to be perpetuated under the government's recent proposals, needs to be looked at again, and urgently. In the emerging pattern, the polytechnics and universities will be more than ever institutions on overlapping spectra.



Whose misconduct?

The delay in performing an investigation of an allegation of misconduct may itself be improper.

THERE must be better ways of conducting an investigation into scientific misconduct than those used by Cornell University Medical Center and the National Institutes of Health (see page 816). This month, five years and four months after its first cursory look at the case, Cornell has quietly confirmed that "an error was present" in a scientific paper published from the laboratory of Dr Jeffrey Borer, a member of the Cornell faculty. Not to be outdone in tardiness, NIH is still working on a final report on the incident, although NIH started slightly later.

The charges against Dr Borer come from his former colleague at Cornell, Dr Jerome Jacobstein, now at Graduate Hospital in Philadelphia. Dr Jacobstein perceives his role as that of "whistle-blower", preventing what he and his lawyer see as a cover-up by Cornell of one of its faculty. Drs Borer and Jacobstein were once co-directors of the laboratory of nuclear cardiology at Cornell. Before their disaffection, they had collaborated on research; Dr Jacobstein would have been a co-author of the disputed paper had these problems not arisen.

After its brief investigation late in 1981, Cornell decided that there was no need to pursue the case; once NIH became involved, it decided it could comfortably leave the investigation to NIH. But NIH botched the job. Early in 1983, it sent to Cornell someone with no medical training; not surprisingly, the first draft of the NIH report appeared to all concerned to be far off the mark. Later, so-called final drafts never appeared to get at the heart of the case. Not until 1985 did NIH add qualified outside experts to its investigation. This expert team apparently finished its report by February 1986, but this was not sent to Cornell until March this year.

Although the case is not a simple one, it is pretty small potatoes as scientific misconduct goes. No one is suggesting that Cornell should forfeit NIH support because some of its research is shoddy, nor is it suggested that Dr Borer should resign in disgrace. No one is even suggesting that the results of the paper in question are invalid, or that the paper should be retracted. The error for which Cornell has concluded Dr Borer is responsible, made not by himself but by a medical student carrying out research in his laboratory, was that of failing to note accurately the number of patients taking a heart drug that could have confused the results of the experiment in question. Given that accurate reporting of the methods and conditions of a scientific study are the *sine qua non* of the literature, there is no doubt that correction or clarification is in order. But one wonders if the pages of journals would not be filled with such clarifications if all those discovered after publication were rushed into print.

To prevent the recurrence of such an error, Cornell will review Dr Borer's supervision of research fellows, publications and research. Dr Borer has agreed to this, as he must, but this is quite a turnabout for an administration that determined five years ago that the case required no further investigation. A more serious charge by Dr Jacobstein is that Dr. Borer "had led a medical student to believe that it was not wrong to misrepresent research methodology". This apparently refers not only to the failure to report the use of confusing drugs, but also to statements about independent scoring of experimental records on early drafts of the paper. Cornell is mute on this charge, except that it agrees to monitor Dr Borer's supervision of student researchers. NIH's opinion will presumably be forthcoming.

The real problem in this case, however, is the fact that it was not settled long ago.

Cornell may have been right to reject Dr Jacobstein's first charges against Dr Borer as being based on a personal dispute, but a more formal inquiry appears to have been called for. Had such an inquiry found Dr Borer to be at fault, steps could have been taken at once that would have ended the case. The same

inquiry could also have determined whether the charges were merely harassment. But when the case first arose, scientific misconduct investigations were still uncharted territory. Now both Cornell and NIH have procedures in place which, they claim, would speed dramatically the resolution of such cases. What the Borer case shows is that speed is essential. Justice that is not swift is often indistinguishable from injustice. □

Technology and terror

Does high technology have a fuller part to play in the control of violence in British Ulster?

THE dilemma of civil authorities faced with domestic insurrection is to balance the steps they take to prevent violence and the cramping of personal liberty entailed in all counter-measures. Sadly, there are all too many parts of the world — the Basque region of Northern Spain, Cyprus (quieter now), the Lebanon and Sri Lanka, for example — in which this question is a live issue. Ulster, the north-west corner of Ireland mostly occupied by the republic of the same name, is a special case for several historical and social reasons, but also because the population affected is both small (5 million altogether) and relatively self-contained. Moreover, the two governments concerned, the British and the Irish, have access to, and the means to deploy, the most technically sophisticated means of security.

In Ireland as elsewhere, the objectives of the security services are to prevent crimes of violence, to which end the identification of those involved in earlier crimes, and of their associates, is a preliminary. (So, too, is the control of explosives and of fire-arms.) What the Ulster experience has so far shown is that traditional policing methods are ineffectual.

The ideal would be an objective means of telling where people are or have been. With such a small population to survey, it is not technically inconceivable that a full-scale Big Brother mechanism could be mounted for keeping track of people, for example by fitting everybody with a personal radio-transmitter in the manner in which some of the more adventurous state jurisdictions in the United States have sought to regulate the movement of criminals on parole. Naturally, the whole of Ireland would be up in arms, and rightly, if such a scheme were seriously suggested, but is it entirely out of court to suggest that the same principle should be applied to motor vehicles?

Ulster's unhappy experience over twenty years is that acts of violence by members of what is called the Irish Republic Army usually involve the use of vehicles. Cars, when packed with high explosives, have proved to be handy mobile bombs. Lord Justice Gibson was killed last week by such a booby trap, a stolen car parked alongside a border roadway. If it were required that all vehicles in Ireland should be fitted with a radio-transmitter merely announcing their true identity to a cellular radio network, it should be much easier than now to tell which vehicles have been stolen and which have carried people to places where terrorists have been at work.

Naturally, for any scheme of this kind to be effective, it would have to be made a criminal offence to disable a transmitter. And while the rules would have to apply to all the vehicles in Ireland, north and south, the security services would not necessarily be overwhelmed by data; mostly, retrospective (and computerized) analysis of the signals gathered by the network would suffice. To safeguard people's rights to use their vehicles for strictly private purposes, it might even be made a condition of the operation of such a scheme that access to the data would be allowed only with judicial consent. No doubt there are other ways in which technology might contribute to the abatement of Ulster terrorism, without which a lasting solution of Ulster's problems is unattainable. Now that the two governments concerned have a mechanism — the Anglo-Irish Council — for dealing with matters of mutual security, there is a case for thinking that they should plan adventurously. □

Beam weapons still science fiction

- US physicists deliver sceptical report
- SDI timetable is threatened

Washington

A DECADE of research is needed even to decide if laser and particle beam weapons are capable of performing the crucial tasks that proponents of the Strategic Defense Initiative (SDI) envisage for them. That is the conclusion of an American Physical Society (APS) report, released last Thursday after a long gestation. In the most thorough examination so far of directed energy weapons, a seventeen-member panel of leading US physicists has concluded that all proposed technologies fall far short of the performance that SDI would demand.

The report has been twenty-one months in the making, the last seven taken up with classification reviews by the Department of Defense (DOD). The panel had access to classified material, so that an authoritative report could be compiled, but public release was then dependent on DOD scrutiny. The amount of material deleted was "minuscule", according to Kumar Patel, research director at Bell Laboratories and co-chairman of the panel with Nicolaas Bloembergen of Harvard University. The most significant omissions in the final report concern the vulnerability of the defensive systems and possible Soviet counter-measures.

Although the report does not address general questions of battle management and computer software complexity, it is much more than a technical assessment of

individual weapons. If the launch of ballistic missiles is accompanied by a much greater number of decoys, the defence has a complex task: it must detect all fast-moving objects (giving only a few false alarms in so doing), distinguish missiles from decoys, track each missile with microradian accuracy, direct an intense, tightly focused beam precisely onto a target, make sure that a kill has been made, and then retarget the beam quickly onto the next missile. In all these facets of operation, and especially in discriminating between missiles and decoys, the report finds many vague concepts and few tested ideas.

SDI research has produced some working devices. A technical summary is in the accompanying inset, but the basic conclusion is that improvement by two orders of magnitude or more in power, for focusing and accuracy is needed for any of today's techniques to find military application. Furthermore, scaling up present devices is much more than a technicality; in many cases fundamental principles are still barely understood.

But the DOD is not put out. In an official response they describe the report as "subjective and unduly pessimistic", and they claim significant progress during the seven months that the report was going through review. They point to a new injector for free-electron lasers at Los Alamos Laboratory, and a high power,

continuous ion source for neutral beam machines, developed at Culham Laboratory in the United Kingdom. However, panel member Andrew Sessler, former director of Lawrence Berkeley Laboratory, said that although these improvements were significant, they concerned only small parts of big devices, and the report's conclusions were by no means invalidated.

The DOD also stresses that it regards a directed energy weapons system as the second phase of SDI, with kinetic energy weapons, which were not studied by the APS panel, constituting the first phase. The APS report puts a decision just on the



What the well dressed California strawberry grower is wearing this season. See p.819.

feasibility of this second phase at least ten years away and, as Jeremiah Sullivan of the University of Illinois emphasized, the whole idea might well turn out to be unworkable. But Lt-Gen. James Abrahamson, of the Strategic Defense Initiative Office, still believes that continued research at current levels will allow a decision to be made in the early 1990s, if not before.

Even if official optimism can overcome the report's sober conclusions, a further revelation contained in the report is bound to increase public and international opposition to SDI. Estimates of the power required for space-based systems show that even for 'standby' operation, each platform must have its own nuclear reactor. The prospects of hundreds of reactors being launched into space cuts deeply into SDI's image as a "non-nuclear" defence shield. Even worse, there is talk of plans to use nuclear explosions in space to power a 'nuclear shotgun'. Hypervelocity pellets powered by a bomb would rip through space, destroying lightweight decoys and leaving warheads to be picked off by other weapons. The system, codenamed Prometheus, is classified, but is thought to have undergone preliminary testing in an underground nuclear explosion.

David Lindley & Alun Anderson

Where the beam-weapon technologies stand:

- **Chemical lasers** based on hydrogen or deuterium fluoride are the most advanced candidate, but their energy output would have to be increased at least 100-fold while maintaining beam quality. Other chemical lasers, such as the atomic iodine type, need to be scaled up by much larger factors.
- **Excimer lasers** fall short by four orders of magnitude in power, and their pulse repetition rate is much too low.
- **Free-electron lasers**, newer and less well understood, are so far almost a million times too feeble.
- **X-ray lasers** pumped by nuclear explosions have been much touted as the answer to SDI's problems, but only a single underground test has been conducted, and the feasibility of a working weapon is entirely undemonstrated.
- **Neutral particle beam weapons** show some promise, but continuous high current beams seem hard to achieve. Neutral particle weapons would have to operate above the atmosphere, and techniques for repeatedly and accurately aiming the beam have yet to be devised.
- **Electron beams** can perhaps be shot along plasma channels punched through the atmosphere by a laser; this prevents the terrestrial

magnetic field from deflecting the beams. But the creation of plasma channels is largely unproved, and the electron beams must be scaled up in voltage, pulse duration and average power.

- **Propagation of intense, tightly focused optical laser beams** presents many problems. A ground-based system would have to allow for atmospheric aberrations, and for the weather (with seven stations in the US, there would on average still be one day a year when all are clouded over). Even in space, novel phase correction methods must be used to keep the beam width down to a diffraction-limited focus.

- **Detection and tracking of missiles** must be almost perfectly reliable. All missiles must be found, and very few false alarms must arise. If missiles are to be tracked during the boost phase by their infrared emission, then ways must be found of locating the body of the missile within the more extensive hot plume. Accurate tracking of missiles during their midcourse, when they are much cooler, is harder still. Long-wavelength infrared telescopes would have to be at least of 10-m aperture to achieve sufficient resolution. Alternative methods, such as microwave or optical radar, would require numerous space platforms.

David Lindley

Cornell misconduct case grinds slowly to its conclusion

Washington

INVESTIGATING scientific misconduct is not easy, a fact underlined by a long, drawn-out investigation of a cardiologist at Cornell University Medical Center in New York that may only now be drawing to a close five and a half years after it was initiated. Cornell has announced disciplinary steps against cardiology researcher Dr Jeffrey S. Borer for scientific practices "below acceptable standards for clinical research". The final chapter in the investigation will come when the National Institutes of Health (NIH) releases a full report on the case.

Cornell's announcement follows the receipt of a final draft of the NIH report. Disciplinary action will include a review of Borer's supervision of research fellows, a review for accuracy of his research papers for a period of at least one year, and a review of a recent published work to make certain that no problems uncovered in the investigation persist.

The doubts about Borer's research stem primarily from a study conducted in late 1981 and ultimately published in the *American Journal of Cardiology* (Jordan *et al.* 51, 1091-1097; 1983). The work looked at cold stress as an alternative to strenuous exercise for inducing ischaemia in patients being examined for coronary artery disease. Dr Jerome G. Jacobstein, then co-director with Borer of the laboratory of nuclear cardiology first raised the allegations of misconduct. Jacobstein's concerns were with the way Borer supervised the medical student conducting the research, and with the accuracy of the language in the methods section of initial drafts of the paper. A committee convened by Cornell in December 1981 dismissed Jacobstein's charges, concluding that no further investigation was warranted. Jacobstein left Cornell in March 1982, for Graduate Hospital in Philadelphia where he is director of nuclear medicine.

But Jacobstein was not satisfied that his charges were being given the attention they deserved. Together with his lawyer, George Washington University law school faculty member Harold Green, Jacobstein took his case to NIH.

In January 1983, a team from NIH visited Cornell to investigate the charges against Borer. After receiving a draft of the initial NIH investigation report, Green additionally charged that some of Borer's subjects had been treated with propranolol within 24 hours of participating in the study, whereas the methods section of the paper stated the opposite. In November 1983, NIH informed Cornell of the new charge, and in March 1984, Cornell acknowledged the reporting error.

Borer informed the *American Journal of Cardiology* of the mistake.

Despite having produced a so-called final report on the case in April 1984, NIH decided to investigate further. In the summer of 1985 NIH called in two outside experts, Harvey Berger, then director of nuclear medicine at Emory University and Bertram Pitt, director of the cardiology division at the University of Michigan, to review the case. Their review was only transmitted to Cornell in March of this year.

NIH will still not comment officially on the case, since its report is not yet completed. NIH is awaiting comments from Cornell and from Jacobstein and Green before declaring the investigation closed. In a prepared statement responding to the Cornell decision, Jacobstein and Green state they are pleased that "after five and a half years of stonewalling" Cornell is taking Jacobstein's allegations seriously. But the statement goes on to criticize Cornell for characterizing Jacobstein's initial allegations as "disagreements", when they were meant to convey the concern that "Dr Borer had led a medical student to believe that it was not wrong to misrepresent research methodology."

Cornell officials do not feel they are acting prematurely in this case even though NIH has not released a final report. Cornell Dean Thomas Meikle is confident that the NIH investigation has uncovered all the relevant facts in this case. For his part, Borer acknowledges responsibility for an error in a study he supervised, and hopes that the book will shortly be closed permanently on the incident.

Joseph Palca

AIDS campaign

Sydney

AUSTRALIA's television campaign to warn of the dangers of AIDS (acquired immune deficiency syndrome) has been a little too successful. People are flocking for blood tests, with a tenfold rise in attendance at some hospitals. But most of them are not in risk groups, and calls to end the campaign are growing.

The campaign, featuring death in the form of the Grim Reaper, was designed to shock. But it did not inform people of the real nature of AIDS, says Professor David Pennington, chairman of an AIDS education committee, AIDS Task Force. The National Advisory Committee on AIDS, who produced the campaign defend their scare tactics. The committee chairperson, Miss Ita Bullrose (herself "radically celibate") says it has "jolted people out of their apathy".

Charles Morgan

New rules for AIDS positives

Washington

US Secretary of Defense Caspar Weinberger last week unveiled new guidelines governing the military's estimated 2,576 personnel who have tested positive for the AIDS (acquired immune deficiency syndrome) virus. So far, 200 out of the 2,100,00 members of the US combined military forces have contracted AIDS.

Under the new Department of Defense (DOD) AIDS policy, those testing positive for AIDS virus (HIV) will be restricted to duty posts within the United States. According to Dr William Mayer, Assistant Secretary of Defense for Health Affairs, the decision will give infected personnel greater access to primary healthcare facilities, and will make it easier to monitor the onset and progress of symptoms. It is also intended to avoid exposing them to risks associated with immunizations required for foreign duty, and to protect them from diseases endemic to foreign countries that their weakened immune systems might not be able to fend off.

The location and types of assignment of HIV-positive service members may also be limited to "protect the health and safety of military personnel with serologic evidence of HIV infection and of other military personnel". Commanders will have the authority to strip them of security clearances and remove them from posts "requiring a high degree of stability or alertness". No decision has been made on whether infected troops will be sent into battle in the event of mobilization.

Infected reservists and students at military academies or under military scholarships are also affected under the new policy. Members of the reserve corps are no longer eligible for military medical treatment if they contract HIV, and students will be expelled at the end of the term or have their scholarships revoked if they become infected.

The guidelines also include provisions for disciplining members of the military who have the AIDS virus, and who knowingly pass it on. Dr Mayer says those who test positive are counselled on "safe sex" procedures, including the use of condoms. Violating those procedures constitutes breaking a direct order, and can lead to criminal prosecution.

The one million civilians who work for DOD will not be subject to routine testing. However, screening civilians for HIV before sending them overseas on military business is being considered by the Pentagon. The State Department now tests members of the Foreign Service, and has just won a case brought against it that alleged that such routine testing violated workers' privacy.

Carol Ezzell

Nakasone on the brink as US trade war threatens

Tokyo

JAPAN'S Prime Minister Yasuhiro Nakasone visits the United States this week having just suffered a humiliating defeat in the Diet over his plans to introduce a sales tax next year. Some commentators are already predicting that Nakasone may be forced to resign after the June summit but much may depend on the outcome of his trip to Washington.

At the end of last year Nakasone was riding the crest of a wave. After winning a landslide victory in the July election, the prime minister forged ahead with the break-up and privatization of the national railways, rewrote party rules and won himself another year in office, and committed Japan to participation in the US Strategic Defense Initiative. But in an attempt to drastically reform Japan's taxation system by introducing a sales tax Nakasone has slammed into a brick wall.

In December a committee of the ruling Liberal Democratic Party (LDP) drew up plans for tax reform that included cuts in income tax, abolition of a tax-free savings system and introduction of a sales tax similar to Europe's VAT (Value Added Tax). From the outset there were howls of protest about the sales tax, not only from the opposition and public but also from rank-and-file members of the LDP.

The opposition parties seized their opportunity and resolutely refused to debate the fiscal 1987 budget in the Diet until the sales tax was scrapped. The sales tax also became the major issue in recent local elections, where Nakasone's plan proved to be a severe liability for the LDP.

But it was in the Diet that Nakasone met defeat. After months of delay the LDP tried to bulldoze the budget through the lower house using sheer weight of numbers (the ruling party holds a clear majority). The opposition responded with a Japanese filibuster — the 'cow walk'. A series of no-confidence motions against key committee members and the finance minister were introduced. Then began a slow-motion voting process. Each member of the opposition walked at a painfully slow pace to the voting podium, stopping after each step to stretch their legs and roll their necks. Voting dragged on day and night as one motion after another was introduced. Finally, last Thursday the bleary-eyed ruling party relented. The sales tax bill was withdrawn and within hours the budget passed the lower house.

This defeat is a major blow for Nakasone, who dug his heels in, in the belief that it was his mission to rehabilitate the nation's finances through tax reform (the government is in debt by more than \$1 million million). In the end it was his

Washington trip that forced Nakasone to retreat.

With US-Japan trade relations in a state of crisis, Nakasone has to present the US with solid evidence that Japan is taking concrete measures to boost domestic demand. But detailed plans of a ¥5 million programme to stimulate the economy could not be drawn up without passage of the budget.

But Nakasone's public image has suffered and failure to ease US-Japan trade friction when in Washington would place him in a very precarious position. On the other hand, if Nakasone can turn his pally relationship with US president Ronald Reagan to advantage and smooth over some of the trade problems he could return home in a greatly strengthened position.

Nakasone aides are already talking of extending the present Diet session into the summer. After discussion with the opposition, the sales tax could be re-introduced in revised form under a different name, and Nakasone is probably the only LDP politician with the nerve to do it.

David Swinbanks

Another UK lab forced to make cuts

London

THE Freshwater Biological Association (FBA), an independent British body that conducts mainly basic research on the ecology of rivers and lakes, may have to shed between 25 and 30 jobs, of a total staff of 103. Two-thirds of FBA's annual income of £2.5 million comes as a grant-in-aid from the government's Natural Environmental Research Council (NERC), with the rest from private contracts. Salary increases in line with Civil Service guidelines, coupled with a reduction in real terms of NERC's contribution and a fall-off in commissioned contracts, mean that a budget deficit of more than £500,000 is projected for next year, with the deficit for the current financial year standing at £340,000.

FBA argues that the basic character of its research is not conducive to success in attracting large numbers of private contracts and it feels it deserves more government support. A final decision on the question of the redundancies will be made next month. Professor Edward Deevey, graduate research curator and professor at the Florida State Museum, described the news as "horrific". "The FBA has a long record of extreme eminence in historical environmental science. A very large number of people are going to be quite upset."

Simon Hadlington

Aqueous award

London

PROFESSOR Tadashi Inagami of Vanderbilt University, Nashville, Tennessee, has been awarded the third Spa Foundation Prize for his work on blood pressure regulators, particularly renin, angiotensin and atrial natriuretic factor. Worth £120,000, the prize was created in 1984 to celebrate the 400th anniversary of the commercial



export of "The Original Spa Water" from the eponymous town in the Ardennes. The Spa prize, administered by Belgium's National Foundation for Scientific Research, is awarded for achievements in research, prevention or therapy of hypertension and cardiovascular disease. C.W.

Hungary for Eureka

London

HUNGARY'S Central Institute for Atmospheric Physics has applied to participate in Eureka, the European Economic Community's programme to foster high technology. The Hungarian move was announced by the West German ministry of research and technology, and the project the Hungarians wish to join is Eurotrac, which aims to monitor the spread of air pollutants in Europe. This is the first formal application from a Comecon country to participate in Eureka. V.R.

Anglo-Dutch merger

GENERAL Electric Company (GEC) of the United Kingdom and Philips of the Netherlands last week announced the merger of their medical equipment divisions to create the world's largest company specializing in medical diagnostic systems. The new company, which will have its headquarters in the United States, is expected to invest particularly heavily in the research and development of computer tomography and magnetic resonance systems.

The merger, between Philips medical systems division and GEC's Picker International subsidiary, will result in annual sales of about \$2,000 million, just ahead of the medical electronics division of General Electric of the United States.

Philips' medical systems division employs 11,000 people with 8,500 based in Europe. Of Picker's workforce of 6,000, 4,000 are based in the United States. GEC will make an additional investment of up to \$200 million in the joint company to bring its share up to 50 per cent. The merger is expected to be completed before the end of the year. S.H.

Can bomb-test swapping cement US – Soviet relationships?

Washington

AN unusual Soviet proposal that the United States and the Soviet Union explode nuclear weapons at each other's test sites has added a new element to the complex negotiations surrounding a test ban treaty. The tests would allow precise calibration of the seismic networks that monitor nuclear blasts. US arms negotiation officials will only describe the new proposals as "interesting".

The inability to monitor accurately nuclear tests has been a major stumbling block in moving towards a nuclear test ban and a source of international tension. The US has repeatedly accused the Soviet Union of violating the 1974 treaty limiting underground tests to a yield of less than 150 kilotons. To circumvent monitoring uncertainties a private US group, the Natural Resources Defense Council (NRDC), last year took the remarkable step of setting up a seismic network inside the Soviet Union, with the cooperation of Soviet scientists. They believed they could show that seismic monitoring really worked. But this development was not welcomed by the United States, which continues to insist that a quite different monitoring technique, continuous reflectometry for radius versus time (CORRTEX) is the only really accurate way to measure the power of nuclear explosions.

CORRTEX is a very much more intrusive monitoring system as it requires a cable to be installed in a borehole drilled just fifty feet from the bomb test chamber. To monitor tests teams of US technicians would have to spend months at Soviet test sites and vice versa: so far the Soviets have opposed the proposal and insisted that seismic monitoring can be adequate. US experts are themselves divided over which technique is best.

Political rather than seismological issues may prompt US insistence on CORRTEX techniques, for the Reagan administration has shown no enthusiasm for a test ban. Continuing nuclear tests are essential to develop the X-ray lasers and other weapons required for the Strategic Defense Initiative (see p.815). And recently nuclear tests have been claimed as necessary to assess the reliability of the nuclear stockpile. According to the US Arms Control and Disarmament Agency, the first priority is reducing the number of nuclear weapons. "If you stop testing you lose reliability and credibility. That leads to a loss of confidence", said an official. This mirrors President Reagan's faith in negotiating from a position of strength.

There is continuing pressure on the Reagan administration to move towards a test ban. On Friday two provisions were

introduced into the House of Representatives that would ban all tests more powerful than one kiloton. Similar legislation was introduced last year and only withdrawn to give Reagan greater freedom of action at his Iceland summit meeting with Soviet Premier Mikhail Gorbachev.

If bomb tests were exchanged both nations would have accurate seismic records of an explosion of known yield. Details of the swap will now be discussed at the meeting of the Nuclear Testing Experts Group at Geneva in early May. But it is thought likely that the US will continue to insist on the CORRTEX techniques that the Soviet Union has so far rejected.

Alun Anderson

• The Soviet Union has been using

underground nuclear explosions for "economic purposes". Peaceful nuclear explosions (PNEs) were at one time advocated for major civil engineering schemes (such as the diversion of rivers) by both US and Soviet experts. The United States, however, soon dropped the idea and since the late 1970s the Soviet Union has said nothing further about using PNEs. On 19 April, however, the Soviet official news agency TASS reported that two underground PNEs of less than 20 ktonnes had been detonated in the area of Perm, and that these had "nothing in common" with the weapon tests two days earlier at the Semipalatinsk test site. On 21 April, the daily *Sovetskaya Rossiya* explained that the Perm explosions had been intended to improve oil recovery by bringing together a series of "mini-formations" by creating a system of fissures. No ecological harm had been done by the explosion, the paper stressed.

Vera Rich

Moscow and Bonn agree to cooperate on nuclear power

Munich

WEST Germany and the Soviet Union have agreed to cooperate on both the scientific and safety aspects of nuclear power. West German Research and Technology Minister Heinz Riesenhuber and the Chairman of the Soviet State Committee for the Utilization of Nuclear Energy, Andronik Petrosyants, signed a five-year pact to this effect in Moscow on 22 April.

The pact, along with agreements on health and agricultural cooperation, brings into operation a 'framework' agreement on scientific and technical cooperation between the two countries that was signed in July 1986. West German Health Minister Rita Süßmuth and her Soviet counterpart Evgenii Chasov signed the health agreement, which provides for the exchange of physicians and medical researchers, on 23 April in Moscow. The agreement on agricultural research is expected to be signed in Bonn during the first half of May.

The details of the nuclear agreement are to be worked out later on this year, reported a speaker at the West German Research Ministry (BMFT). The BMFT's draft protocol includes plans for co-operation on gas-, water- and sodium-cooled reactors, nuclear fusion and materials research as well as research into treatment of radioactive waste.

The pact also provides for industrial co-operation from the German side. West Germany's leading producer of nuclear power plants, Kraftwerk Union (KWU), is eager to sell service contracts to Soviet plants. Unlike the scientific exchange, which is to be supported financially by both sides, these contracts are to be paid

for by the Soviets in the form of cash or goods, said KWU spokesman Wolfgang Breyer. "KWU can play a significant role" in increasing the standards of safety in Soviet power plants, said Breyer, "assuming that the Soviet Union wants our help."

The German-Soviet agreement reflects the significant warming of relations between the two countries in 1987. But it also promises to provide useful scientific information to both sides. "We want to see what the Soviets have learned from running sodium-cooled reactors over long periods," said Gerhard Heusener of the German Nuclear Research Centre in Karlsruhe. Heusener explained that the Soviets use sodium-cooled reactors like the BN350 to provide energy for desalination plants, which in turn provide drinking water, and therefore the Soviets need a high level of reliability.

The negotiations leading up to the new agreement had been plagued with difficulties. For years, discussion foundered on the status of West German research institutions in West Berlin. As the Soviets refuse to recognize the right of such institutions to be situated in Berlin, still officially administered by the four victorious powers in the Second World War, such researchers can only be listed by name in the pact, without their respective institutions. This so-called *ad personam* solution was reached in 1984.

The news of the nuclear agreement was accompanied by Soviet Nuclear Energy Minister Nikolai Lukonin's announcement on 22 April in Moscow that the Soviets plan to increase threefold the amount of electricity produced by nuclear power by 1995.

Steven Dickman

Polish government getting the message on environment

London

ENVIRONMENTAL issues have provided much controversy in Poland in recent years — notably the continued emission of sulphur dioxide at a rate of over 4 million tonnes annually from coal-burning power stations. Now, in what will be seen by the government's opponents as an attempt to prevent radical political groups from making ground by exploiting 'green' issues, the government has established an Ecological Social Movement to coordinate the work of clubs, committees and individuals for the protection of the environment and cultural monuments.

The new movement, under the chairmanship of the president of the Polish Academy of Sciences, Professor Jan Kostrzewski, will formally be a "self-governing and independent body", working under the patronage of PRON, the Patriotic Committee for National Rebirth. PRON was set up by the government during the martial law period to channel the various initiatives launched by the outlawed Solidarity trade union.

During the past year the extra-legal movement Freedom and Peace has won considerable public sympathy by espousing 'green' issues. Freedom and Peace

a purely practical point of view, some central coordination of endeavours is necessary. Even one of the government's own programmes, that for environmental education in schools and universities, is being implemented unevenly and would benefit from better control, according to a recent study by Lublin Polytechnic. Another argument for a central body is that it may give more emphasis to the problems of the more remote areas, away from the large conurbations where pressure groups are more active. Perhaps the danger to Europe's last remaining tract of ancient forest, the Białowieża, would have been more widely known if

the new movement had been in existence earlier. The forest, straddling the Soviet — Polish border, may suffer irreparable harm when ground-water levels increase with the opening of the Siemionówka dam on the Narew river.

But a major reason for the foundation of the Ecological Social Movement may be a foreign policy move. A Polish — Swedish cooperation programme on acid rain has been under discussion for some time and is at last taking shape. The Swedish negotiators, however, feel that progress would be much quicker if they were dealing with non-governmental organizations, rather than direct with government. The Swedes have suggested the Kraków Ecological Club as a suitable partner, but the new movement would provide a suitable organization operating at the national level.

Vera Rich

Field test of ice-minus bacteria goes ahead despite vandals

San Francisco

AFTER a prolonged legal battle and a last-minute sabotage attempt, field tests of genetically engineered ice-minus bacteria have finally begun in California. Advanced Genetic Sciences (AGS) sprayed the bacteria, designed to protect plants from frost damage, on a strawberry patch in Contra Costa county at dawn on 24 April, the day after receiving the go-ahead from a Superior Court judge.

When the AGS scientists arrived at the site at 5 a.m., they found that vandals had slipped past two security guards, entered the fenced lot and uprooted 80 per cent of the stawberry plants. Most of the plants were intact, however, and once they had been replanted, Julie Lindemann, project leader, began applying the bacteria. Company officials said that the vandalism would not significantly disrupt the experiment.

The project has suffered many legal delays. Plans for tests in Monterey County last year were blocked by local residents. Last spring the Environmental Protection Agency (EPA) suspended AGS's experimental permit and fined the company \$13,000 for conducting unapproved pathogenicity tests on fruit trees on the roof of their building. The EPA reinstated approval in September, and in February AGS chose the Contra Costa location. Contra Costa residents at first showed little concern, but opposition was raised by environmental groups from Berkeley and Monterey.

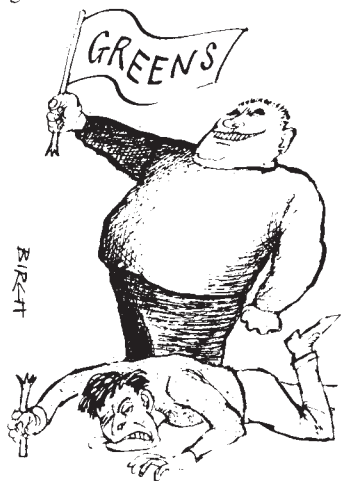
The test was approved by the California Department of Food and Agriculture (CDFA) on 13 April, but was delayed further by a suit filed by several environmental groups, including Jeremy Rifkin's Foundation on Economic Trends.

CDFA approval was based on hundreds of tests showing the bacteria will not spread to untreated plants and are not pathogenic to animals or to 47 different species of native and crop plants. The bacteria, which AGS plans to market under the trade name Frostban, are a mixture of two strains of *Pseudomonas* that lack the gene coding for an ice nucleation protein. They are classified as a pesticide because they minimize colonization of leaves and blossoms by ice-forming *Pseudomonas* strains.

Lindemann applied the bacteria with a hand pump sprayer that produces large droplets, minimizing aerosol formation. She wore a protective suit and face mask required by the CDFA for testing of unregistered pesticides. Flats of barley plants in the 0.8 acres of barren buffer zone which surrounds the 0.2 acre strawberry patch will act as "traps" to monitor for the escape of bacteria, and air samplers will be used to check for airborne bacteria leaving the plot.

The goal of the field tests is to show whether Frostban can hold its own against naturally occurring bacteria and provide effective frost protection. During the six week experiment, researchers will take plants back to the laboratory to count bacteria, and will also measure the temperature at which the blossoms freeze. Unfortunately many blossoms were destroyed by the vandals, so some of the experiments will have to wait until after the second Frostban application in 10 days. Lindemann expects plenty of new blossoms by then. Dozens more field tests will be necessary, before AGS can market the product, a goal they hope to realize in three years, said AGS vice president, John Bedbrook.

Marcia Barinaga



began as a campaign by young people against compulsory military service, objecting particularly to the military oath in which draftees express loyalty to the Soviet Union. Even the respectable Kraków Ecological Club has spoken up for the more radical Freedom and Peace group, suggesting that the Wrocław city authorities should pay the fines imposed on those who took part in demonstrations against the Wrocław steel mills.

With its official links, the Ecological Social Movement will not be able to include the Freedom and Peace campaigners. But it would be simplistic to view the new body simply as a government ploy to 'hijack' the ecological movement. From

NASA seeks to move some space science missions off shuttle . . .

Munich/Washington

THE National Aeronautics and Space Administration (NASA) is planning to move several science missions now scheduled for launch aboard the space shuttle to expendable launch vehicles (ELVs). One beneficiary of the new plan will be ROSAT, a German X-ray satellite that was supposed to be put into orbit by the US space shuttle this year. Joachim Trümper, director of the Max Planck Institute for Extraterrestrial Physics in Garching, West Germany where ROSAT was developed, says he is operating on the tentative assumption that ROSAT will be launched in 1989 aboard a Delta 2 rocket.

NASA appears to be heeding the cry of many in the space science community to seek ELVs as an alternative to the shuttle. A NASA Advisory Council report released on 12 March urged NASA to seek a mixed fleet for future missions. In addition to ROSAT, NASA now plans to use a Delta 2 to launch two other astrophysics satellites, the extended ultraviolet explorer and the cosmic background explorer. NASA has already found money in its budget to buy the Delta 2 for the cosmic background explorer mission, and is seeking permission from Congress to reallocate \$2 million from its 1987 budget for ELVs. An item for ELVs may be added for 1988 if the White House Office of Management and Budget approves.

NASA also plans to purchase two Titan 4 rockets. One will be a back-up to the shuttle for either the Galileo mission to Jupiter or the joint Ulysses mission with the European Space Agency to explore the Sun's poles. The other Titan 4 would be used to launch a tracking and data relay satellite now scheduled for a shuttle launch in the 1990s.

The backlog in shuttle payloads following the Challenger accident would probably have delayed ROSAT until 1994. NASA's decision to use Delta 2s as launch vehicles follows the Air Force decision to use the same rocket as its primary medium-launch vehicle. ROSAT is just inches too broad to fit on top of a regular Delta 2, and a new shroud will have to be designed, at an increased cost of several million dollars. ROSAT programme managers would have preferred using an Atlas Centaur because it is about three inches wider and would not have required any such modification.

There is still a chance that an Atlas Centaur rocket may become available for ROSAT. The National Ocean and Atmospheric Administration (NOAA) is said to be considering buying several Atlas Centaur rockets to launch future weather satellites. This could bring the price down on the Atlas Centaur to make it more competitive with the less expensive Delta 2.

Steven Dickman & Joseph Palca

. . . and military give ground

Washington/London/Tokyo

WARRING factions within the Reagan administration have resolved their differences over how to proceed with negotiations between the United States and its international partners on the manned space station. Disputes had erupted between the National Aeronautics and Space Administration (NASA) and the Defense Department over the use of the space station for military activities.

Negotiations will resume this week in Ottawa with the Canadian space station participants. On 4 May talks resume with Japan, and on 11 May representatives of the European Space Agency (ESA) are due in Washington.

The Pentagon agreed to tone down its demand that the space station be available for possible national security activities following an interagency meeting at the White House on 17 April. A letter from Secretary of Defense Caspar Weinberger to Secretary of State George Shultz had urged that the United States explicitly reserve the right to "conduct national security activities on the US elements of the Space Station, without the approval or

review of other nations". The US position now emphasizes that the space station will be used for "peaceful purposes consistent with international law."

ESA has welcomed this as "a positive step forward". But Dr Wilhelm Brado, head of ESA's cabinet, said there were still some problems to be resolved. Dr Brado says each partner in the space station project will have to decide whether its own national plans comply with the concept of peaceful purposes.

Some of the problems have been left to be sorted out at the agency level in the working agreement between NASA and ESA, according to Dr Brado. The Japanese government, concerned about possible military use of the space station, is being cautious, however. Mr Masahiro Kawasaki, deputy director general of Research and Development at the Science and Technology Agency says that they will be waiting to hear the details at the meeting with the US negotiators next week. "Japan's position has been and still is that the space station should only be used for peaceful purposes", he stressed.

Late last week NASA told contractors

Isis economizes on electricity

London

THE Isis pulsed-neutron facility at the Rutherford Appleton laboratory in Oxfordshire is to close down for six months from September, largely because it cannot afford to pay the electricity bill. Lack of government grants to the Science and Engineering Research Council (SERC), which funds Isis, has resulted in a £2 million shortfall for the current financial year.

The long-term future of Isis, known until its official inauguration in October 1985 as the spallation neutron source (SNS), now appears to depend on putative European partners jointly funding the facility. In December 1985, memoranda of understanding for such a joint enterprise were signed by France and Italy, and later by Sweden. Interest has been shown by the Netherlands and West Germany. Negotiations are continuing towards an agreement on the constitution of a consortium and levels of funding.

Last year's SERC grant to Isis was £9.97 million, with a further £830,000 earned from other sources. This year the grant was reduced to £8.5 million, with a projected income from other sources, including instrumentation from India, Japan and Italy, totalling £1.5 million.

This year's shutdown period, which will include between two and three months for upgrading, will be when electricity tariffs are at their highest. When in full operation, Isis is closed down for two or three months each year for routine maintenance.

The extended closure will have implications for Britain's contribution to the understanding of how high-temperature superconductors work. Dr Donald Paul, of Warwick University, who has been using Isis for neutron-scattering studies on superconductors, says Britain will now start to lag behind its competitors in the rapidly moving field. "It's undoubtedly frustrating," Dr Paul says.

Simon Hadlington

to prepare their bids on various portions of the space station. ESA is expecting to be ready to receive bids once its programme is refined and presented to its ministers in November.

The Defense Department requested late last year that negotiations with international partners be suspended until Pentagon interests could be considered. Once talks resumed, it became clear that military activities aboard the space station could present problems for some countries, and the debate among US agencies has been over how to retain the full range of US interests without jeopardizing international cooperation.

Kathy Johnston, Joseph Palca & David Swinbanks

Nuclear power yes, Chernobyl reactor technology no

London

A YEAR after the Chernobyl disaster, the Soviet Union remains firmly committed to nuclear power, both for electrical energy and district heating. But the RBMK channel-type uranium-graphite reactor, the type used at Chernobyl, seems to be on its way out. During the past two weeks, Soviet officials have indicated that the "RBMK technology is outdated" and will not be used for further power stations after eight now under construction are completed. Furthermore, according to the Swedish daily *Upsala Nya Tidning*, Evgenii Larin, an official at the Soviet Ministry of Nuclear Energy, has stated that the fourth reactor at Ignalina, an RBMK-based power station in eastern Lithuania, will now not be completed.

Since Chernobyl, both Sweden and Finland have been worried about the proximity of the RBMK stations at Ignalina and Leningrad; indeed, the Finnish government at one point made overtures to Moscow, suggesting that Finland would be willing to bear the cost of constructing Western-type containment buildings at the two sites, arguing that even if as the Soviets claim such structures are, objectively, unnecessary, they would be worth the expense to reassure the Finnish public.

Ignalina is situated in the Lithuanian lake district, an area of considerable agricultural and recreational importance. Last summer, shortly after the Chernobyl disaster, a former environment official at Ignalina, Dr Kazys Eringis, who now lives in the United States, warned that sloppy construction and operating techniques prevalent at the station could mean the emission of significant radioactive contaminants and, in the case of an accident, could render the whole of Lithuania, Latvia and Byelorussia uninhabitable. Lithuanian sources indicate that there has been considerable opposition to the expansion of Ignalina, not least because (as with a similar anti-nuclear movement in Soviet Armenia) the republic already has ample supplies of electrical energy for its own need, and the further expansion of the nuclear programme is simply to 'export' energy to other parts of the Soviet Union.

According to Dr Valerii Legasov, the Deputy Director of the Kurchatov Nuclear Power Institute in Moscow, and the chief Soviet spokesman at the Post-accident Review meeting on Chernobyl in Vienna last August, Soviet plans have long envisaged the phasing out of the RBMK, in favour of the safer and more modern water-water reactor, as soon as the giant nuclear equipment enterprise

Atommasch could supply the necessary machinery. Atommasch, however, was the focus of a major scandal four years ago, when a trouble-shooter from Moscow had to be sent to sort out inefficiency and corruption, and, according to Legasov, only the first stage of the plant is operating.

With the phasing out of the RBMK, the Soviet energy programme will be committed, in the short term, to the VVER, although long-term hopes and research programmes are focused on fast breeders (two are already in operation) and, eventually, on fusion power. Meanwhile the official view is that Chernobyl was the result of a breakdown of "work discipline", which led to unauthorized experiments with the reactor and more efficient training programmes, stricter operating rules and the modification of equipment to make unauthorized tampering physically impossible will be enough to guarantee safety.

Nevertheless, some experiments with power stations still continue within the socialist bloc. *Izvestiya* recently reported experiments at the Dukovany power station in Czechoslovakia in which the hydraulic tests on the reactor and the circulatory flushing of the primary circuit during the tuning and start-up operations were carried out simultaneously instead of, as previously, successively. The new procedure was said to have the approval of top-level experts, and to be not only faster (a saving of 66 days was reported) but also safer. This may well be so, but in the aftermath of Chernobyl any reports of reactor experiments, however highly approved, will undoubtedly arouse some public anxieties.

Vera Rich

UK upland grazing still contaminated

London

A YEAR after Chernobyl, tests are still under way to determine levels of caesium contamination in sheep in northern England and Wales, as government restrictions on the movement and slaughter of the animals remain in force. Initial mathematical model predictions of a quick decline in caesium levels proved incorrect because they failed to consider the special soil conditions in the uplands; the acidic peat soils with low clay content do not bind the caesium, which enters plant tissue to be consumed by the sheep. Over 270,000 sheep, plus their lambs, are under restrictions, compared to 4.2 million initially.

The British government has so far paid out £4.3 million in compensation to 7,500 farmers. As the radioactivity persists, officials have not yet been able to tell farmers when the restrictions on their live-stock might be lifted, or even whether next year's lambs will be free from contamination. Live monitoring techniques have shown ten per cent of the restricted animals tested since January have caesium levels over 1,000 becquerels per kilogram, the EEC-agreed limit for foodstuffs which is far more stringent than the pre-Chernobyl limit recommended by the National Radiological Protection Board of 9,000 becquerels per kilogram.

Researchers at Bristol University are assessing radionuclide levels in grass and soil samples from uplands areas, by creating spring-like conditions to promote plant growth. The movement of caesium from the soil to plants and animals is being studied by the Institute for Terrestrial Ecology, which is aiming to establish the characteristics of areas that can absorb contamination.

Kathy Johnston



Jeff Golblum (right) plays Jim Watson to Tim Piggott-Smith's Francis Crick in the BBC's *Horizon* special, 'Life Story', first broadcast on BBC2 this week.

Postdocs need a career structure

SIR—The recent letter (*Nature* 325, 478; 1987) from D. J. Murphy, who suggests removing age-related postdoc salary scales in British universities, offers a shallow solution to a grave problem, and begs the question of what level the postdoc salary would be set at. We presume, as the aim of such an exercise is to allow older postdocs to compete with their younger counterparts, that they are deemed to be worth less than their current salaries, and not that younger postdocs are underpaid. But reducing the salary of all postdocs to the £8,000–£9,000 level would mean that some older postdocs would be paid less than the technical staff with whom they work.

Murphy says this solution might not entirely halt the increasing drain of talent away from Britain. We suggest it would most certainly increase the drain. Not only that, it would drive many talented scientists out of science altogether.

Murphy also fails to appreciate that many older postdocs find themselves in their present position not by choice but simply because there is minimal recruitment of lecturers and government research scientists. Older postdocs do not want to be postdocs for the rest of their lives, to act as “assistants” (*Nature* 326, 122; 1987) to older scientists who are fortunate enough to be tenured. They are certainly not prepared to do this for the level of salary Murphy envisages.

Older postdocs are certainly a “formidable pool of talent for the nation” but let us not devise schemes to further exploit them. The “obvious solution” is to provide them with employment as independent scientists and hence permit realization of their worth.

J.R. SPEAKMAN
G.R. IASON
M.S. GRISLEY

University of Aberdeen,
Department of Zoology,
Aberdeen AB9 2TN, UK

SIR—P. N. Campbell (*Nature* 326, 122; 1987) highlights the plight of postdoctoral research assistants. In response to an advertisement for a research fellowship at a northern technological university, I was appointed as a research assistant as I had yet to complete my PhD. I had no argument with this decision. Much to the embarrassment of my professor, however, the university refused to promote me to research fellow upon completion of my PhD, appointing me instead to postdoctoral research assistant. I then had to reapply for my own job (and be reinterviewed by four senior academics) before being appointed as a research fellow. During the interview, my professor refused to ask me any questions, as he con-

sidered university administrators and outmoded regulations were to blame for the situation. If universities can be responsible for such time-wasting farces, then it is no wonder the government is scrutinizing their activities.

N.H. DAY

39 Bowmans Avenue, Hitchin,
Hertfordshire SG4 9QD, UK

First things first

SIR—The question of whether authors should use the word ‘first’ to imply a priority research finding has received considerable comment in *Nature* recently^{1,2}. It is of course demanding for a referee or editor to be vigilant and review the associated literature to ensure that such statements are warranted, should editorial policy permit them. However, it was surprising to read in *Nature* of the “first experimental data for prokaryotic regulation at distances >1kb”³ when in a recent review published in *Nature* references to work⁴ which could have drawn the same priority statement were made, as indeed might have a letter published last March⁵. Perhaps *Nature* should adopt the same policy² as *Proceedings of the National Academy of Sciences of the USA* with respect to priority statements to protect not only the authors but the journal also.

MARTIN BUCK
RAY DIXON

MARTIN DRUMMOND

ARC Unit of Nitrogen Fixation,
University of Sussex,
Brighton BN1 9RQ, UK

1. Maddox, J. *Nature* 324, 509 (1986)
2. Zwanzig, F.R. *Nature* 325, 754 (1987)
3. Dandanell, G. *et al.* *Nature* 325, 823 (1987)
4. Ptashne, M. *Nature* 322, 697 (1986)
5. Reitzer, L.J. & Magasanik, B. *Cell* 45, 785 (1986)
6. Buck, M. *et al.* *Nature* 320, 374 (1986)

Fundamental split

SIR—The recent reports in *Nature* concerning the steady decline in the quality and quantity of British science and the lack of cohesion between fundamental and applied research have been illuminating but have often lacked concrete examples of political and professional decisions that will have an adverse effect on an important area of science in which Britain currently has a toe-hold. The recent decisions to sell the breeding programmes at the Plant Breeding Institute (PBI) and to relocate the remaining public sector scientists at the John Innes Institute and Rothamsted Experimental Station are clear examples of political shortsightedness that will critically weaken an important sector of British science.

The immediate effect of these decisions will be to drive skilled researchers who

wish to work in a secure environment from the Agricultural and Food Research Council (AFRC).

Second, the decisions will separate the disciplines of molecular biology and plant breeding which many of us at the PBI have been striving to join. The benefits of these collaborations have become apparent recently in planned field trials of genetically engineered potatoes and the use of DNA probes in breeding programmes, and it is generally believed that the application of these methods of molecular biology will have a profound effect on plant breeding and agriculture. Indeed, many chemical companies interested in genetic engineering have purchased seed companies to obtain breeding teams. It will take many years to recreate the special interaction between breeder and molecular biologist that exists at the PBI; in fact we do not know if the AFRC intends to continue this work. Given the parlous state of the economy, I doubt if extra money will be provided for breeding programmes at the John Innes and Rothamsted.

Finally, moving the molecular biology activities of the PBI from Cambridge will remove the stimulating day-to-day contact we have with biologists at the university and the Laboratory of Molecular Biology. It may also prove more difficult to attract visiting scientists who are at present drawn by the special scientific atmosphere of Cambridge.

MICHAEL BEVAN

Plant Breeding Institute,
Maris Lane,
Cambridge CB2 2LQ, UK

Colonizing space

SIR—Robert Shepherd's letter on the colonization of space (*Nature* 326, 124; 1987) misses an important point. It is true that an attempt to colonize the other planets in the Solar System while they are in their present state would probably be unproductive, but it may be possible to change their environments (at least for Mars and Venus) so that they become habitable¹⁻³. Moreover, it seems rather unlikely that extra-solar planets will have environments that are perfectly suitable for human habitation, and colonization may require that their original environments be changed. Such a procedure would greatly increase the number of habitable planets in the Galaxy, and experience gained in our own Solar System would be invaluable.

IAN CRAWFORD

University of London Observatory
Annexe and Planetary Image Centre,
33-35 Daws Lane,
London NW7 4SD, UK

1. Zwicky, F. *Observatory* 68, 121 (1948).
2. Sagan, C. *Science* 133, 849 (1961).
3. Allaby, M. & Lovelock, J. *The Greening of Mars* (Deutsch, London, 1984).

Amorphous ice still a puzzle

The idea that there is more than one form of amorphous ice has been around for the past three years, but explanation of the true state of affairs is still elusive.

AMORPHOUS ice must be the strangest stuff. Although the first reports that ordinary water can be solidified in a non-crystalline form came from observations modelled on classical lines of solidified supercooled water and the deposition of water vapour on cold surfaces, the bulk production of amorphous ice was first reported only in 1984, when Mishima, Calvert and Whalley at the National Research Council in Ottawa (*Nature* **310**, 393; 1984) described experiments in which ice cooled to 77 K was subjected to pressures greater than 10 kbar.

The first surprise was the formation of a material much denser than expected (1.31 g cm^{-3}), which was shown by X-ray diffraction to be structurally disordered. The second, that the material passed through a succession of amorphous phases as the pressure was allowed to relax to normal; in the extreme case, the material could be returned to zero pressure (still at 77 K) in its high-density amorphous form. A year later (*Nature* **314**, 76; 1985) the same group described experiments which appeared to show that first-order phase transitions (where there is a finite latent heat) can actually occur between different amorphous phases. Now two theoreticians, John S. Tse and Michael L. Klein, also from Ottawa, have been able to throw some light on this strange behaviour (*Phys. Rev. Lett.* **58**, 1672; 1987).

That the phase diagram of ice is like a wayward piece of crazy paving is well known. Ice I, the form of solid water stable at room temperature and pressure, is but one of eight or nine crystalline forms to have been identified, although not always fully characterized. But the widely understood property of ice I, that its melting point decreases with increasing pressure (which is why skates laden with the weight of a human form move smoothly over ordinary ice), does not apply to all the forms of solid crystalline ice that can be directly in equilibrium with liquid water: ices III, VI and VII, and possibly ice IV as well, resemble more common materials in that their melting points increase with pressure (which is the same as saying that the specific volume of the liquid is greater than that of the solid). The other known forms of ice appear as transformations at temperatures well below the melting point of ice I at any pressure, and must be likened to the phase transitions of mantle rocks induced within the Earth's crust by pressure alone.

Mishima, Calvert and Whalley were originally concerned with the question of what would happen to ordinary ice at a low temperature when it was squeezed by a pressure great enough to turn it into something else. And what would that something else be? At zero temperature and infinite pressure, there is no difficulty in guessing that ice should be a crystalline structure in which the oxygen atoms are arranged in a body-centred cubic lattice. But at higher temperatures or at pressures less than infinity, there is room for disordered structures — liquids or, the same thing, the solid liquids called glasses.

So how is it possible to understand the transition from crystalline ice I to amorphous ice? Tse and Klein rely heavily on the original observation that there appear to be two amorphous states of ice, one with a density of 1.31 g cm^{-3} and the other with a (zero-pressure) density of 0.91 g cm^{-3} . They note that what X-ray diffraction shows is that the two amorphous forms of ice differ not so much in the separation of the nearest-neighbour oxygen atoms from each other, but in the distribution of the next-nearest neighbours. In the low-density form, the next-nearest neighbour separation is much like that in ordinary ice, $4.5 \text{ \AA}$ or thereabouts, but there is a peak at around $3.5 \text{ \AA}$ in the high-pressure form. Given the observation by Mishima *et al.* in 1985 that the two distinct phases transform one into the other with the production of measurable latent heat, it is reasonable to expect that there will a simple model to describe each of the two amorphous phases.

That may be asking a lot, for Tse and Klein have used the technique of molecular dynamics, which entails the calculation of the motion of all the atoms in a molecular system of some kind. Inevitably, there are physical limits to what even the largest machines can accomplish, so that the various versions of solid ice have been represented by groups of 128 water molecules, with appropriate boundary conditions which ensure that their calculated properties are repeated elsewhere in the structure of the entire solid. Inevitably, there are imprecisions; phase transitions turn out to be less sharp than those observed, for example.

That ice I is a tetrahedral structure, in which each oxygen is surrounded symmetrically by four others, is nevertheless confirmed; that well-established experi-

mental observation has long since been verified by calculation. The question of what happens to ice I (at low temperature) as the pressure is increased has been pursued by means of an enormous expenditure of computer power, yielding among other things a pairwise correlation function for oxygen atoms suggesting that there is a range of pressures — not too great and not too small — in which the interstices between neighbouring oxygen tetrahedra are progressively filled with other oxygen atoms, each carrying its two protons. This is the explanation of the dramatic 30 per cent decrease of specific volume at the transition between ice I and high-density disordered ice. Tse and Klein have less specific things to say about the low-density amorphous phase, although they are able to report agreement between the radial distribution functions and those inferred from experiment.

There are all kinds of hidden benefits in calculations of this kind. For example, the molecular dynamics method should allow its users to say something about the positions of the hydrogen atoms in the squeezed states of ice (on the average, they will be midway between the oxygen atoms), which are not of course observable; that means that it should be possible to say something about the specific entropy of the various systems, which must matter a great deal in the high-pressure phases. But the real worry about such calculations at very high pressure is that the specific free energy is great enough for questions of molecular stability to be relevant; could it even be that some of these shadowy amorphous phases differ from each other in the way in which electrically polarized microstructures impinge on each other? Certainly it would be of more than passing interest if the calculations could be made to yield predictions of measurable electrical properties.

People, it appears, will have to make the best they can of interatomic distance distribution functions whose peaks do not even unambiguously count the frequency of the occasions of which a next-nearest neighbour is separated from the centre not by a nearest neighbour but by nothing at all. There was always a hope that the discovery of an apparent first-order transition between one amorphous form of ice and another would have been accounted for by a simple structural transformation, but that has been disappointed.

John Maddox

Cystic fibrosis

Classical and reverse genetics

Peter N. Goodfellow

EVERYONE KNOWS a carrier of the mutant cystic fibrosis (CF) gene. The disease is manifest, however, only when two carriers reproduce and a child inherits the mutant gene from both parents. Much effort has been directed towards elucidating the molecular genetics of the disease, and two approaches now seem to be producing hopeful results. In one, Dorin *et al.*, who reported their results in *Nature* earlier this month¹, use a 'classical' approach to isolate the CF antigen gene. In the other, Estivill *et al.*, who report their findings on page 840 of this issue², use a new method that may identify the region of chromosome 7 containing the CF gene.

Almost 1 in every 20 caucasians is a carrier of CF, and about 1 in 2,000 children born in Britain suffer from the disease. Usually, the most obvious symptom in the afflicted child is the overproduction of mucus, which results in poor clearance of the airways with subsequent bacterial colonization and lung damage. It is only recently, because of steady improvements in antibiotic treatment and physiotherapy, that the life expectancy of the CF child has extended into adulthood.

Working backwards from the clinical symptoms, biochemists and cellular physiologists have found defective chloride transport in CF epithelial cells (see the recent discussion in *News and Views*³). One consequence of the defect seems to be the overproduction of a normal serum protein, the CF antigen⁴. It should be stressed that mutation in the CF antigen gene is not the cause of CF, as the CF gene is on chromosome 7 whereas the CF antigen gene is on chromosome 1 (ref. 5). By purifying the protein, sequencing the amino terminus and making a highly redundant oligomeric DNA probe, Dorin *et al.* have isolated complementary DNA clones encoding the CF antigen¹. The deduced amino-acid sequence of the whole CF antigen reveals extensive similarity to several calcium-binding proteins. One possibility, hinted at by Dorin *et al.*, is that the CF antigen is part of a macromolecular signal-transducing complex coupled to the chloride channel. The CF gene product could be another component of the complex.

Isolation of the CF antigen gene illustrates the classical approach for cloning mammalian genes of medical and biological interest. In an attempt to clone the CF gene Estivill *et al.* in this issue² use the alternative approach of 'reverse' genetics. This approach initially ignores biochem-

ical clues from the phenotype and relies instead on the chromosomal localization of the disease gene (see my recent *News and Views* article⁶ for a review). These methods have yielded spectacular success for several disease genes and CF may soon be added to the list.

Previous experiments have already localized the CF gene close to the proto-oncogene *met* (ref. 7) and the locus *D7S8* (ref. 8), defined by the probe *J3.11*. This region of the human genome has been isolated in a mouse cell by using the chromosome fragmentation technique of chromosome-mediated gene transfer (CMGT). One of the CMGT transfectants, CIId, includes *met*, and contains only 10⁶ base pairs of human DNA⁹. The problem was how to identify the CF gene in this chromosomal fragment.

Bird has proposed the provocative hypothesis that many genes can be identified by the presence of associated unmethylated CpG-rich regions called HTF islands¹⁰. These regions can be recognized as the sites of cleavage by methylation-sensitive enzymes such as *HpaII* (hence *HpaII* tiny fragments or HTF) and most of the rare-cutting restriction enzymes used for long-range restriction mapping. The recent work of Lindsay and Bird¹¹ supports this hypothesis, but it is not known what proportion of genes lack HTF islands or how many islands lack genes.

Starting with DNA from the transfectant CIId, Estivill *et al.* used a new cosmid vector designed to clone genomic DNA fragments with one end derived from cleavage at unmethylated *XmaIII* sites. Such sites should be at least partially diagnostic for HTF islands. After screening the library with total human genomic DNA, Estivill *et al.* obtained several cosmids that contain DNA derived from human chromosome 7. These cosmids were used to isolate further sequences and to define an HTF island that maps between *met* and *D7S8*.

High levels of linkage disequilibrium between CF and polymorphic loci associated with the HTF island suggest that the island must be close to the CF gene. The unpublished data of Estivill *et al.* demonstrate a transcript associated with the island, and molecular analysis of cystic fibrosis sufferers will confirm or refute the hope that this is the CF gene. Of immediate importance, the strong linkage disequilibrium will help to identify CF carriers. Molecular genetics continues to bring good news to those families afflicted with genetic disease. □

1. Dorin, J. *et al.* *Nature* **326**, 614-617 (1987).
2. Estivill, X. *et al.* *Nature* **326**, 840-845 (1987).
3. Case, M. *Nature* **322**, 407 (1986).
4. Bullock, S., Hayward, C., Manson, J. C., Brock, D. J. H. & Raeburn, J. A. *Clin. Genet.* **21**, 336-341 (1982).
5. van Heyningen, V., Hayward, C., Fletcher, J. & McAuley, C. *Nature* **315**, 513-515 (1985).
6. Goodfellow, P. N. *Nature* **322**, 12 (1986).
7. White, R. *et al.* *Nature* **318**, 382-384 (1985).
8. Wainwright, B. J. *et al.* *Nature* **318**, 384-385 (1985).
9. Scambler, P. J. *et al.* *Nucleic Acids Res.* **14**, 7159-7174 (1986).
10. Bird, A. P. *Nature* **321**, 209-213 (1986).
11. Lindsay, S. & Bird, A. P. *Nature* (in the press).

Peter N. Goodfellow is at the Imperial Cancer Research Fund Laboratories, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

Parkinson's disease

A cure using brain transplants?

Solomon H. Snyder

PATIENTS with neuronal disorders of the brain are rarely, if ever, cured, because neurons of the adult central nervous system do not regenerate. For years neuroscientists have talked about nervous tissue transplants as the ultimate solution for paralysis following spinal cord damage, tissue loss following trauma to the brain, strokes, tumours or degenerative disabilities such as Alzheimer's disease and Parkinson's disease. This quest has spawned abundant research into factors that might permit neurons of the adult central nervous system to divide, just as they do during fetal life. Many groups have isolated neural growth factors with possible physiological roles. Nerve growth factor, well known as a survival factor for peripheral sympathetic nerves, now appears also to promote survival of the septo-hippo-

campal acetylcholine-containing neuronal pathway whose function is disrupted in Alzheimer's disease¹. Now, Madrazo and associates in Mexico City report the relief of parkinsonian symptoms², in severely afflicted, human patients receiving autografts of their adrenal medulla to the caudate nucleus on one side of the brain. Their published report describes only two patients but, according to press reports, they have had similar success with eight subsequent patients.

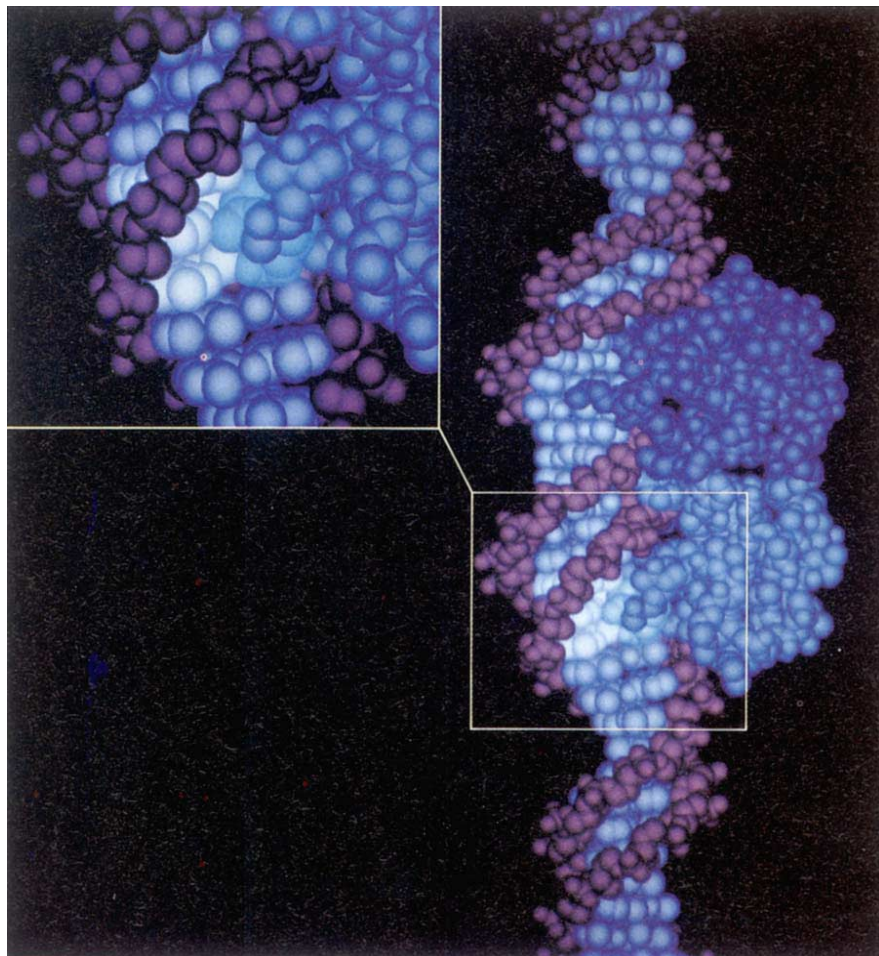
Parkinson's disease has been a particular focus of research because animal models are available. The major symptoms of idiopathic Parkinson's disease appear attributable to degeneration of the principal dopamine pathway of the brain, whose cell bodies are in the substantia nigra of the brain stem with axons

ascending to terminate in the caudate and putamen, areas of the brain involved in the control of motor activity. Replacement of the missing dopamine by treatment with its amino-acid precursor L-Dopa (L-dihydroxyphenylalanine) is highly effective therapy, although it does not halt the progressive deterioration of patients. Unilateral destruction of the nigrostriatal dopamine pathway causes rats literally to rotate in their cages. Implants into the denervated caudate of a substantia nigra from a fetal rat sprout dopamine-secreting processes from the graft, whereupon the locomotor abnormalities are relieved³. Because the adrenal medulla secretes catecholamines including dopamine, adrenal transplants to lesioned caudates of rats give similar improvement⁴.

One of the patients of Madrazo *et al.*² displayed almost total relief of symptoms for 3 months and the other for 10 months, implying that the graft has permanently ameliorated the neurological abnormality, providing a genuine 'cure'. Before making such a strong assertion, a few caveats should be emphasized. The initial patients were somewhat atypical, being quite young and with extremely severe symptoms. The sister of one of the patients also had severe parkinsonism at an early age, suggesting a genetic aetiology, which is also atypical.

Some cases of Parkinson's disease are caused by toxins such as manganese or drugs such as MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine). Although there is no direct evidence for drug or toxin exposure in any of the patients, this possibility is hard to exclude. Regardless of the aetiology of subtypes of Parkinson's disease, however, the chief neuropathological abnormality, a loss of the nigrostriatal dopamine pathway, is almost invariably the same. Thus, a treatment for one class of patients should be applicable to all.

Some aspects of the new work² suggest specific mechanisms to account for the therapeutic success. The Mexican surgeons implanted the adrenal tissue onto the surface of the caudate nucleus in close contact with ventricular fluid. In a previous unsuccessful study, Swedish neurosurgeons grafted the adrenal medulla directly into the caudate parenchyma⁵. Conceivably, contiguity with the ventricular fluid provides access to essential nutrients. Alternatively, the ventricular fluid may be a way of transporting dopamine or some other factor from the transplant to other sites. Interestingly, the Mexican transplants are unilateral but motor improvement of the patients is bilateral. This result excludes neuronal-like sprouting from the graft as a therapeutic mechanism, because such sprouting would only be unilateral. Moore⁶, reviewing the Mexican work, suggested that what was transplanted might be a growth



The picture shows the repressor protein of bacteriophage 434 bound to synthetic operator DNA, with the enlarged detail highlighting interactions between repressor side-chains (light blue) and operator base pairs (white) that contribute to the sequence-specificity of binding. The structure of this complex, obtained by X-ray crystallography, is reported by J.E. Anderson, M. Ptashne and S.C. Harrison on pages 846–852 of this issue, and provides the first detailed view of a gene-regulatory protein bound to DNA. One surprising feature of the structure is that the DNA helix at the centre of the operator is overwound, and the significance of this is strengthened by the results of G.B. Koudelka, Harrison and Ptashne reported on pages 886–888 of this issue, showing that only certain base pairs can be accommodated at the central positions of the operator without decreasing the affinity for repressor, even though these sites appear not to be directly contacted by any repressor side-chains. The implication is that these central base pairs contribute to the specificity of binding by their influence on the conformation of the operator DNA helix. □

factor that would influence the caudate on both sides of the brain. If this is the case, one might attempt to isolate such a factor. Because adrenal grafts relieve nigrostriatal lesion deficits in rats, there is a readily available animal model to test purified factors.

The Mexican work has implications for brain transplants in other diseases. A model of Alzheimer's disease in rats, in which lesions of the septo-hippocampal cholinergic pathway produces learning and memory abnormalities, is 'cured' both behaviourally and neurochemically by transplants of fetal septal tissue⁷. If, as suggested by the Mexican neurosurgeons, the adrenal grafts are acting as biological infusion pumps for dopamine, one might consider grafting into the brains of

Alzheimer's patients some acetylcholine-producing peripheral tissue from the same patient rather than attempting the virtually impossible task of securing septal tissue from a human fetus. Similarly, relatively non-specific transplants might be considered in the treatment of spinal cord or brain trauma, stroke and brain tumours. □

1. Hefti, F. & Weiner, U.J. *Ann. Neurol.* **20**, 275-281 (1986).
2. Madrazo, I. *et al. New. Engl. J. Med.* **316**, 831-834 (1987).
3. Bjorklund, A., Dunnet, S.B., Stenevi, U., Lewis, M.E. & Iversen, S.D. *Brain Res.* **199**, 307-333 (1980).
4. Freed, W.J. *Biol. Psychiat.* **18**, 1205-1267 (1983).
5. Backlund, E.-O. *et al. J. Neurosurg.* **62**, 169-173 (1985).
6. Moore, R.Y. *New. Engl. J. Med.* **316**, 872-873 (1987).
7. Bjorklund, A. & Stenevi, U. *A. Rev. Neurosci.* **7**, 279-308 (1984).

Solomon H. Snyder is in the Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA.

Topology

The area of the plane

Ian Stewart

WHAT is the area of the plane? I am thinking of the whole infinite plane here, not just a part of it, so obviously the area is infinite. But if the question is phrased more carefully, it leads to a fascinating minimization problem, which has recently been solved by Christophe Bavard and Pierre Pansu of the Ecole Polytechnique (*Annales Scientifiques de l'Ecole Normale Supérieure* **19**, 479–490; 1986). The answer, by the way, is $2\pi(1+\sqrt{2})$. The trick is that there are more ways to define distances in the plane, and hence more ways to calculate areas, than just the usual one for which the area is infinite.

Problems of this type have a lengthy pedigree, going back to the tale of Queen Dido and the bull's hide. What is the greatest area of land that can be enclosed by a bull's hide? Answer: cut the hide into a long string, and then enclose a circle — the largest area with a given perimeter. Such questions are called variational problems, and more sophisticated versions underlie some of the deepest parts of mathematical physics.

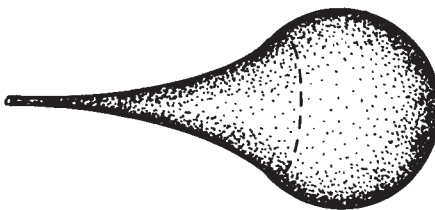
The problem is easier to grasp if instead of a plane we consider a sphere. The usual way to measure the distance between two points on a sphere is to join them by a great circle, and measure along that. If the radius of the sphere is r then the area corresponding to this notion of distance is the usual area $4\pi r^2$. This depends on r . But there is another fundamental property of the sphere that also depends on r , its curvature, which is $1/r$. Suppose that we insist that the curvature be less than or equal to 1. Then r is greater than or equal to 1, so the area is greater than or equal to 4π . Thus, for standard spheres, if the curvature is less than or equal to 1 then the minimal area is 4π . And it turns out that a similar statement is true, even if we use a notion of distance on the sphere that is not the usual one.

The proper mathematical setting for studying such questions is the theory of riemannian manifolds: multidimensional analogues of surfaces, equipped with a riemannian metric, or 'small-scale distance structure'. The idea goes back to Bernhard Riemann in a famous lecture of 1854, and forms the basis of Einstein's theory of general relativity. There are two different ways to think of such structures, which might be called the extrinsic and intrinsic viewpoints.

Imagine a perfect sphere in ordinary three-dimensional space. Consider also the same sphere, but distorted into a much less perfect egg. Define distances on both surfaces by measuring along geodesics,

shortest paths. Then the distance between two points on the original sphere will change as the sphere turns into an egg. Further, the change will vary with the amount of distortion. Where the sphere gets stretched a lot, on the flatter parts of the egg, distances expand; where it gets shrunk, near the pointy bit, they contract.

The extrinsic view of this is that there is a single distance structure on the surrounding three-space, but two distinct surfaces, sphere and egg. The intrinsic view is that there is the same surface, a sphere, but two distinct distance structures: the usual one, and the one that it inherits if it is thought of as being egg-



To form the 'plane' of minimal area, join a sphere of curvature 1 to a pseudosphere (the trumpet-like surface on the right) of curvature -1 . The combined surface has curvature between -1 and 1 at every point, total area $4\pi(1+\sqrt{2})$, and is topologically equivalent to the plane. Any other space equivalent to the plane, with curvature between -1 and 1 , must have a larger area.

shaped. It was Riemann himself who urged the importance of the intrinsic view, because it allows one to deal with the surface in its own right, rather than having to imagine some larger surrounding space.

The crucial idea of curvature is most familiar in the extrinsic picture. We can immediately see that the pointed end of an egg is more sharply curved than the flatter sides. But, in 1827, Carl Friedrich Gauss found a way to think of the curvature intrinsically, and Riemann extended this approach to multidimensional manifolds. The basic idea is quite simple. Pick a point on the surface, and draw round it a circle of small radius r . If the circumference of the circle is $2\pi r$, as in the plane, the curvature at that point is 0. If the circumference is less than $2\pi r$, the curvature is positive; if greater, the curvature is negative.

This makes good sense. If you take a circular disk and shrink its edge (say by tightening a draw-string running round it) it will crumple up into a hollow cup, which is curved like a sphere. If you expand the perimeter, say by taking a disk of polythene and stretching its edge here and there, it will not lie flat, nor will it stretch over a sphere. Instead you have to put pleats in

the edge to flatten it. But it will stretch over a suitably curved saddle shape, and this is what negatively curved surfaces look like. So, at a point of positive curvature a surface looks like a spherical bowl; at a point of negative curvature it looks like a saddle; and at a point of zero curvature it looks flat. Moreover, you can decide which, without thinking of the surface sitting in any external space.

Riemann considered spaces whose curvature varies from one point to another: positive here, negative there, zero somewhere else. Einstein exploited this idea of 'curved space' in his theory of gravitation. Roughly speaking, near a massive body, space is more highly curved. Light and particles follow geodesics. In highly curved regions geodesics become bent, so light and particles travel along curved paths. But this notion of curvature reflects not so much an actual bending of the space, as a distortion of its metric, away from the metric appropriate to flat space. The simplest spaces are those with constant curvature. A sphere, with its usual metric, has constant positive curvature. A plane, with its usual metric, has constant zero curvature. One example of a surface of constant negative curvature is the pseudosphere, shaped rather like a trumpet.

M. Gromov (*Publications Mathématiques de l'IHES* **56**, 5–100; 1982) introduced the idea of the minimal volume of an n -dimensional riemannian manifold. This is the smallest value of its n -dimensional volume, calculated for all possible metrics, such that at all points the curvature lies between -1 and $+1$. For a surface, the two-dimensional 'volume' is the area, so we must find the minimal area.

The possible topological types of surface have long been known. For example, if there is no boundary and the surface is closed and two-sided, it must be a sphere with g handles stuck on. By the classical Gauss–Bonnet formula, it follows that the minimal area is then $2\pi|2g-2|$. A similar result holds for one-sided surfaces, and surfaces with edges. For a cylinder or Möbius band the minimal area is 0. The only difficult case is when the surface is topologically the same as a plane.

Gromov proved that the minimal area of a plane is at least $4\pi+0.01$. This estimate was improved by Bavard (*J. différent. Geom.* **19**, 137–142; 1984) to $4\pi+0.484$. Gromov conjectured that the correct answer is $2\pi(1+\sqrt{2})$, a value achieved by the teardrop-shaped combination of sphere and pseudosphere illustrated in the figure. Note that this surface, topologically, really is just a plane: if it is opened out along the narrow tube leading off to infinity, it can be stretched flat.

The problem, then, boils down to showing that any other metric on the plane, of curvature between -1 and $+1$, leads to an area that is larger than $2\pi(1+\sqrt{2})$. It is

this that Bavard and Pansu prove. Considering the 'elementary' nature of the problem it requires surprisingly sophisticated methods, including the theory of 'currents' introduced by W. Fleming (*Trans. Am. math. Soc.* **121**, 160–168; 1966).

Gromov's imaginative idea of the minimal volume of a space produces a new topological invariant: here a number is associated with each topological space. New invariants are always interesting because

they suggest new problems, related in unexpected ways to existing invariants and have the power to open up new lines of attack on old problems. But an invariant is useful only if you can calculate it. Bavard and Pansu fill an irritating gap in our knowledge of surfaces, and in so doing they remind us that simple questions can have non-trivial, elegant answers. □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Superconducting ceramics

The electrons are paired

David Caplin

THE crucial step in the understanding of superconductivity came in 1956 with Cooper's idea that, unlike a normal metal where the conduction electrons are more or less independent of each other, in a superconductor formation of electron pairs reduces the energy of the system. Most physicists have thought it likely that the new superconductors, which are made from mixtures of metal oxides and have transition temperatures five or ten times higher than conventional superconductors, also contain Cooper pairs, but now firm experimental proof has come with an elegant experiment by Colin Gough and colleagues at Birmingham University, reported on page 855 of this issue.

The Birmingham experiment depends on the response of a superconductor to a magnetic field. The classical Lorentz force on a charged particle moving in a magnetic field translates quantum mechanically into a relationship between the field and the quantum-mechanical phase θ . In a superconducting ring threaded by magnetic flux Φ , the phase of the wavefunction that describes the superconducting electrons advances steadily around the ring, and accumulates an increment of $2\pi q\Phi/h$ for each circuit, where h is Planck's constant and q the charge of the carriers. The wavefunction at any point on the ring has to be single valued, so one circuit of the ring must correspond to an exact integral number of phase cycles, $\theta=2n\pi$, forcing the contained flux to be quantized in units of h/q . For Cooper

pairs, with $q=2e$, the superconducting flux-quantum is $h/2e$, denoted by Φ_0 . The magnitude of Φ_0 is 2.07×10^{-15} Wb, roughly the amount of flux from the Earth's magnetic field (typically 10^{-4} T) that threads a 5- μm -diameter loop.

The first direct demonstration of flux quantization in superconductors was by Deaver and Fairbank (Deaver, B.S. & Fairbank, W.M., *Phys. Rev. Lett.* **7**, 43–46; 1961). They measured the magnetic moment of tin cylinders, diameter ~ 20 μm , made by electroplating a few micrometres of tin onto a fine, copper wire. As the external magnetic field was increased, a superconducting screening current was set up in the tin cylinder to keep the contained flux constant; when some critical field was exceeded, the contained flux jumped to a new value. Several such steps could be seen, and agreed with the theoretical value of Φ_0 within ~ 20 per cent.

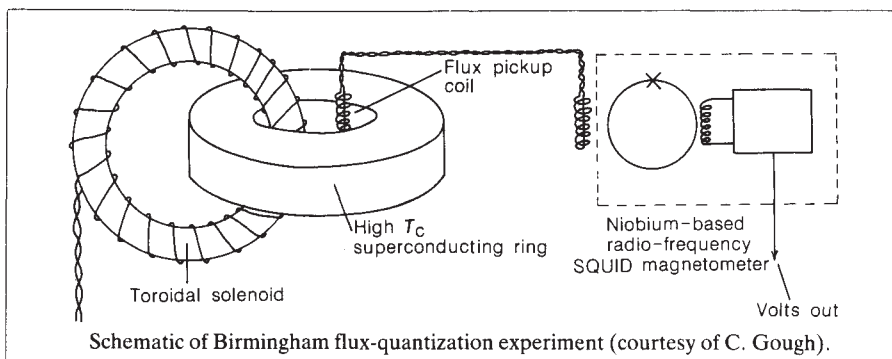
The principle of the Birmingham experiment is the same, but it differs in detail. Although thin films of the new materials have been prepared and are superconducting (see, for example, Somekh *et al.* on page 857 of this issue), it is unlikely that a direct repetition of the Deaver–Fairbank experiment, using a thin-walled sample, would display flux quantization. In their present form, these materials consist of microscopic particles that have been sintered together; contacts between the particles are only weakly superconducting (similar 'weak links' in conventional superconductors are well understood).

Hence they can carry only low current densities, and substantial thicknesses of material are needed for the screening currents to be fully effective against changes in external field.

The Birmingham ring, made of sintered oxides to give composition $\text{Y}_{1-x}\text{Ba}_x\text{CuO}_4$, is therefore rather chunky, with a 10-mm external and 4.5-mm internal diameter. The large cross-sectional area means that the magnetic field corresponding to a flux quantum is small, $\sim 10^{-10}$ T, so that the experiment has to be well screened from external-field disturbances, and also a sensitive detector for the magnetization in the ring is needed. Both requirements can be met quite readily using conventional superconductors: a lead shield, superconducting in liquid helium, excludes external-field changes, and a commercial SQUID (superconducting quantum interference device) sensor monitors the magnetic response of the ring.

Even so, the experiment would be difficult to do simply by applying an external field and looking for steps in the response. It is the magnetic flux — not the field — that is quantized, and one needs to know the effective cross-sectional area of the ring to translate field to flux. That area is somewhere between the areas corresponding to the inner and outer diameters, and depends on the detailed distribution of supercurrents within the ring. The Birmingham group have ingeniously circumvented the effective area problem. They link a toroidal solenoid, wound from ordinary copper wire, with the ring, so that the solenoid and the ring look like two links in a chain as shown in the figure. The field and flux inside a toroidal solenoid can be calculated using elementary electromagnetism, and there is no field outside the solenoid. Thus, a known current through the solenoid injects a known amount of flux into the interior of the superconducting ring. The ring, being superconducting, responds by setting up a screening current to ensure that the total contained flux is unchanged. The magnitude of the screening current is monitored with another small coil connected to the SQUID sensor. With this arrangement, as the solenoid current is increased, the screening current and the SQUID output increase in sympathy so long as the ring stays perfectly superconducting.

In practice, because of the weak links in the ring, the critical current is soon exceeded and the enclosed flux jumps back by some number of quanta; as the solenoid current is increased further, the process repeats itself and generates a saw-tooth output-pattern. The size of the flux quantum can be extracted from the magnitude of these jumps, but it turns out that they correspond typically to a hundred or so quanta, making accurate estimation difficult. Smaller jumps are needed, and in the second part of the experiment,



with zero solenoid current, the superconducting ring is 'kicked', by exposing the ring to bursts of electrical interference, between states containing slightly different numbers of flux quanta (this magnetic flux comes from whatever residual magnetic field has been trapped inside the lead shield). The associated changes in screening current and SQUID output can then be translated into magnetic flux using a calibration derived from the first part of the experiment; the key point is that accurate knowledge of the effective cross-sectional area of the ring is unnecessary. The result for the magnitude of the flux quantum, $0.97 \pm 0.04 \Phi_0$, is conclusive

evidence that in the new materials, as in conventional superconductors, the electrons are paired.

It would have been very unsettling had the new materials not contained electron pairs. With that reassurance, the question now is whether the Bardeen-Cooper-Shrieffer pairing mechanism, involving interaction of the conduction electrons with the quantized vibrations of the ions in the lattice, is responsible for the high transition temperatures, or if some other, stronger mechanism must be invoked. □

David Caplin is in the Department of Physics, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BZ, UK.

Palaeontology

This is the forest primaeval

Christopher J. Cleal

IT WOULD be unfair to criticize Longfellow for describing, in his epic poem *Evangeline*, the present-day Nova Scotian forests as primaeval. There is a mute darkness about them, readily evoking the beginnings of time. The 300-million-year-old Carboniferous equatorial swamps are, however, the true primaeval forests. They were the earliest really extensive development of dense afforestation, spreading over large areas of what is now Europe and North America, and they contained little terrestrial animal life other than insects, spiders and the occasional amphibian. Individual fossil plants from these forests have been studied for more than 200 years, and many are now very well understood; but what is less clear is the community structure of the forests themselves. This is no mere academic exercise, as these forests produced major deposits of coals and hydrocarbon gases. Its potential to assist in the exploitation of these fossil fuels is part of the reason for an upsurge in interest in establishing the community structure of these forests, progress in which is reported in this issue by R.A. Gastaldo (*Nature* **326**, 869–871; 1987).

Early published reconstructions of the forests were rather fanciful, simply showing all the plants of the Carboniferous fossil record growing together in the same swampy habitat. During the past couple of decades, however, palaeontologists have produced improved reconstructions, by comparing the fossils with the sedimentary environments in which they are found. Probably the best-known studies of this type are by A.C. Scott (*Proc. Yorks. geol. Soc.* **41**, 461–508; 1978 and *Proc. geol. Ass.* **90**, 97–116; 1979), who has produced a model for the British Coal Measures swamps with distinctive flood-basin, lakeside and levee-top assemblages. The difficulty with this type of study is that it is mainly based on compression

fossils preserved in mudstones and sandstones, representing plant fragments which were transported by wind and/or water away from where they originally grew. That a fossil is found in, say, a flood-basin deposit does not prove that it came from a plant growing in that environment. By analysing structures in the sediments that indicate directions of transport, some of these difficulties can be overcome but, without evidence from *in situ* fossils, uncertainty remains about the model.

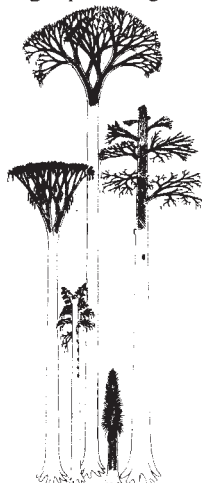
An alternative approach is to examine the plant fragments in the coalified peats which, it is believed, were preserved nearer to where they grew. An early attempt was A.H.V. Smith's palynological study of the Coal Measures coal seams (*Proc. Yorks. geol. Soc.* **33**, 432–474; 1962), in which he identified several different habitats. The difficulty here is the uncertainty both as to which plants produced the various types of pollen and spores, and as to the extent of drift of the pollen and spores, which introduces considerable noise into the results. A more robust technique is the examination of larger plant organs (seeds, branches) from

the coalified peats, which are easier to identify with the parent plant, and which are less subject to dispersal. Normally, the peats are too compressed to see the individual component plant fragments but, if the peat has been impregnated by calcareous minerals before compression to form 'coal balls', the plant fragments can be directly observed and identified. A recent analysis of the distribution of plants in some North American coal balls by W.A. DiMichele and T.L. Phillips (see figure) shows several different assemblages within the peat-forming habitat, including pioneer, dominant and opportunist communities; and, until recently, represented the most refined model available for these forests.

Nevertheless, such coal-ball studies cannot answer all our questions about the ecology of these forests. They are still based on plant fragments which may have been subject to some post-mortem transportation. Also, they can tell us only about the structure of the peat-forming swamp communities, and not the other habitats where peats did not form, such as the levees and the lakesides. Gastaldo's report in this issue overcomes some of these limitations. His distributional analysis of *in situ* tree stumps and forest litter exposed in some coal workings in Alabama allows him to refine some of the ideas of DiMichele and Phillips. The wettest parts of the swamps seem to have been occupied by arborescent club-mosses of the genus *Lepidophloios*, but became replaced in the rather drier areas by stands of another genus of large club-moss, *Lepidodendron*. Unfortunately, he finds no evidence of the *Anabathra*-dominated pioneer community recognized by DiMichele and Phillips. Gastaldo could show the gradation into the drier levee habitat, however, of interest because the plants growing on the levees produced the bulk of the compression fossils, including those used for the stratigraphical analysis of these strata; and previously their exact relationship to the swamp floras was not properly understood. Gastaldo's study improves the robustness of such stratigraphical work.

Gastaldo argues that accurate reconstructions of these forests can be produced only by the study of such *in situ* fossils. This is perhaps a rather exaggerated view, and the other approaches will no doubt continue to play an important role. Nevertheless, he is undoubtedly correct that his type of work provides the most reliable data for interpreting community structures within fossil forests. *Evangeline* and Gabriel may never have wandered through these ancient woodlands, but at least we now have a better idea of what it would be like to have been there. □

Christopher J. Cleal is a freelance palaeontologist at 37 Lipscombe Close, Newbury, Berkshire RG14 5JW, UK.



Reconstructions of some of the major arborescent lycophytes that grew in the Coal Measures forests. From left to right: *Lepidophloios hali*, *Anabathra pulcherrima*, *Diaphrodendron dicentricum*, an immature tree of general form, and *Diaphrodendron scleroticum*. (From DiMichele, W.A. & Phillips, T.L. *Rev. Palaeobot. Palynol.* **44**, 1–26; 1985.)

100 years ago

PRACTICAL ELECTRICITY

Practical Electricity. By W. E. Ayrton, F.R.S. (London: Cassell and Co., 1887.)

Prof. Ayrton's book on *Practical Electricity* fills a gap, and is well fitted for the purpose for which its author has designed it. The book comprises the substance of the first year's course for students of electrical technology in the City and Guilds Central Institution at South Kensington. Starting from the definition of an ampere, the various means of measuring currents are described, and full details are given of the methods for carrying out the experiments, for recording results graphically and otherwise, and for drawing conclusions from the experiments. The illustrations show clearly the arrangement of apparatus in each case. In Prof. Ayrton's laboratory the apparatus requires for any one series of observations is mounted permanently on the same board; the student finds everything ready and the necessary connexions made when he begins. Thus, Fig. 15 shows the apparatus for investigating the action of a current on a magnet. The large coil CC is so arranged that the current can

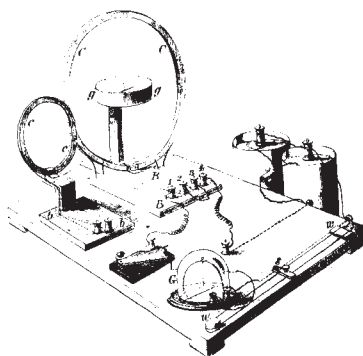


FIG. 15

be made to traverse it eight, twelve or sixteen times, and its effect on the magnet *gg* observed; or the coil *CC* can be replaced by *cc* of half the radius, which has four turns on it; or again, *CC* can be used simultaneously with *cc*, the current being sent in opposite directions through the two. *ww* is a resistance by means of which the current may be varied at will, and *G* a galvanometer.

From *Nature* 35, 601; 28 April 1887.

Cosmology

Where is Population III?

Bernard Carr

A RECENT paper by Cayrel¹ adds an interesting twist to speculation on the whereabouts of those most elusive of astronomical objects, Population III stars. These are the stars of zero metallicity which must have formed before the Population I stars in galactic disks and Population II stars in galactic halos. The existence of such stars is inevitable because metals can be made only by stars themselves (unlike some of the lighter elements, such as helium and deuterium, which are produced through cosmological nucleosynthesis in the Big Bang). Thus, the first stars must have contained no metals at all. One might expect Population III stars to have the same sort of distribution of masses as stars forming today, in which case some should be small enough (smaller than $0.8 M_{\odot}$, where $M_{\odot}$ is the mass of the Sun) still to be burning their nuclear fuel. The problem is that, despite extensive searches, nobody has ever found a zero-metallicity star. Bond², for example, has surveyed an appreciable fraction of the whole sky and yet found only two stars with a metallicity Z as small as $10^{-3} Z_{\odot}$ and none with Z as small as $10^{-4} Z_{\odot}$. ($Z_{\odot}$ is the ratio of the mass in iron to the mass in hydrogen in the Sun, $\sim 10^{-2}$.)

One possible explanation is to assume that all the first stars were more massive than $0.8 M_{\odot}$, but not so massive that they evolved to black holes without returning any heavy elements to the background medium. In this case, they would produce an initial burst of enrichment, thereby imposing a minimum metallicity $Z_{\min}$ for

all subsequent stars, without surviving until the present epoch. But why should $Z_{\min}$ be $\sim 10^{-5}$? About 10 years ago, before the absence of metal-free stars was so conspicuous, Hartquist and Cameron³ invoked a possible feedback mechanism that explains this. They argued that the first burst of star formation should stop once the stars have produced enough photons to ionize the Universe. This happens when $\sim 10^{-5}$ of the mass in gas has been converted into stars, which — for a reasonable metal yield — gives the sort of value for $Z_{\min}$ required.

Cayrel's paper¹, with the rather curious title "And if Population III were Population II?", proposes a different kind of feedback mechanism that seems to have rather more power of explanation. He starts with the suggestion of Fall and Rees⁴ that the first protogalactic clouds probably had a mass in the range 10^4 – $10^7 M_{\odot}$. He then envisages the following evolutionary sequence: first, the core of each cloud collapses to form a compact cluster of massive O stars. (This is not implausible as this is the way large clouds seem to evolve in the present epoch; for example, the central object R136a in the 30 Doradus complex of the Large Magellanic Cloud is thought to contain several dozen O stars.) Because massive stars suffer considerable mass loss during their nuclear burning phase and ultimately explode as supernovae, one expects those in the core of the cluster to generate a strong outflow of material. A shock will be produced where this material encounters gas falling

inwards in the outer envelope of the cloud. So Cayrel envisages, as a second step in the evolutionary process, another generation of stars forming in the dense shell of gas behind the shock. These stars should be smaller than the first ones because cooling (and hence fragmentation) will be facilitated both by the high density within the shell and by the presence of the metals produced by the supernovae in the core. Thus, a crucial feature of this model is that the Population III clouds produce both massive short-lived stars (that generate an initial enrichment) and Population II type, long-lived stars (whose abundance reflects that enrichment).

The suggestion that Population III clouds could produce a bimodal mass spectrum of stellar fragments is not new. In a somewhat different context, for example, Kashlinsky and Rees⁵ proposed that the first clouds to bind in the 'hierarchical clustering model' for cosmic structure, should evolve to centrifugally supported disks of low-mass stars, each with a single massive star at its centre. Cayrel's picture, however, has an important additional feature in that the explosions which generate the metals also blast away an appreciable fraction of the cloud mass, partly in the form of stars and partly as unprocessed gas. The idea that systems with sufficiently small binding energy can lose mass as a result of supernova activity is already quite popular; in particular, Dekel and Silk⁶ have argued that this applies for systems smaller than $10^9 M_{\odot}$, the characteristic mass of a dwarf galaxy. If one invokes the same idea in the Cayrel context, one is naturally led to make the following identifications: the surviving cluster of stars with a globular cluster; the ejected stars with field Population II objects; and the ejected gas with the material that eventually condenses into galactic disks (Population I stars).

It is the suggestion that Population II stars are made in the original Population III clouds that motivates the title of Cayrel's paper. Actually, his insistence on calling the second-generation stars Population III is rather puzzling as, were his proposal ever confirmed, it would be much more sensible to reserve the term Population III for the massive stars which form in the cluster cores — they do represent, after all, a distinct population and they are the only zero-metallicity objects. Semantics aside, the model is attractive because it makes specific quantitative predictions. For example, if the original cloud has a mass of $10^6 M_{\odot}$ and a radius of 1 pc (as expected), then one knows that each supernova explosion releases enough energy to free $\sim 10^3 M_{\odot}$ of the material in each cloud. As one expects one supernova for every $100 M_{\odot}$ of gas and a metal yield of about 10 per cent from each exploding star, this corresponds to a metal enrich-

ment $Z \sim 0.1 Z_{\odot}$, comparable to that in globular clusters. The model also predicts that the ratio of the mass in halo stars to that in disk stars should be ~ 0.1 (as observed) as this corresponds to the efficiency with which gas is processed into stars in the original clouds.

The proposal also links up nicely with recent suggestions by Zinn⁷ that there are two populations of globular clusters: one distributed in a spheroid with a metallicity peaking at $\sim 0.03 Z_{\odot}$, the other in a thick disk with a higher (though still subsolar) metallicity peaking at $\sim 0.3 Z_{\odot}$. Cayrel suggests that the first consists of the surviving primordial clusters, whereas the second is formed later and has a higher metallicity as it is further contaminated with metals generated by stars made in the galactic spheroid. The fact that the metals in the original stars were generated by massive stars alone may also explain specific abundance anomalies associated with metal-poor objects. In particular the relatively high abundances of oxygen and nitrogen compared with iron.

Whether or not Cayrel's suggestion turns out to be viable, his paper is important because it focuses attention on the crucial but unsettled question of what happens to those first bound objects that, in nearly all theories for the origin of large-scale structure, must have been formed before galaxies themselves. Whether these objects appeared well before galaxies (as in the hierarchical clustering model) or at about the same time (as in the Fall-Rees model presupposed by Cayrel), the issues raised are very similar. What sort of stars do they form (if any) and what sort of astrophysical feedback do they have on subsequent galactic or cosmological evolution? There is little consensus, as yet, on either of these issues. Cayrel's picture might be regarded as a compromise between those more extreme models that either neglect Population III, or attribute to it even more exotic roles. $\square$

1. Cayrel, R. *Astron. Astrophys.* **168**, 81–88 (1987).
2. Bond, H.E. *Astrophys. J.* **248**, 606–611 (1981).
3. Hartquist, T.W. & Cameron, A.G.W. *Astr. Sp. Sci.* **48**, 145–158 (1987).
4. Fall, S.M. & Rees, M.J. *Astrophys. J.* **298**, 18–26 (1985).
5. Kashlinsky, A. & Rees, M.J. *Mon. Not. R. astr. Soc.* **205**, 955–971 (1983).
6. Dekel, A. & Silk, J. *Astrophys. J.* **303**, 39–55 (1986).
7. Zinn, R. *Astrophys. J.* **293**, 424–444 (1985).

Bernard Carr is at the School of Mathematical Sciences, Queen Mary College, University of London, London E1 4NS, UK.

Erratum

In the article "Microtubule activation of kinesin ATPase activity" by Stanley Cohn (*Nature* **326**, 16–17; 1987) the figure was taken from ref. 8 (M. Porter et al. *J. biol. Chem.* **262**, 2794–2802; 1987). In the last paragraph of Dr Cohen's article, the value cited of 10–20 μM for the half-maximal velocity of kinesin-induced microtubule gliding was found with sea urchin egg kinesin⁸. The value for squid neuron kinesin is higher, about 60–100 μM . $\square$

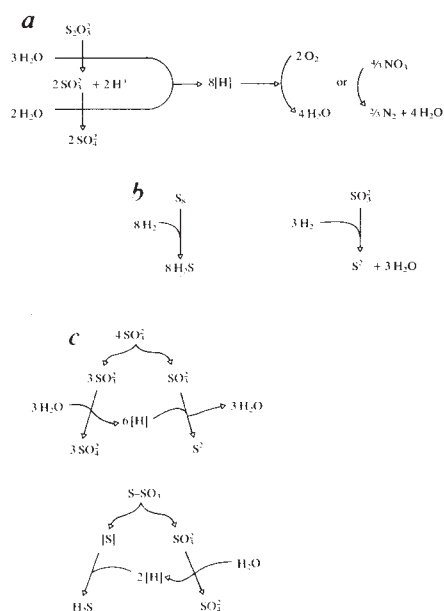
Inorganic fermentation

Sulphur bacteria first again

Donovan P. Kelly

ON PAGE 891 of this issue¹ Friedhelm Bak and Heribert Cypionka describe how some anaerobic sulphur bacteria can obtain energy for growth by the fermentation of inorganic sulphur compounds in a manner that has never been described before. Their observations open up a new field of study in microbial metabolism, and alter some fundamental and long-held views about the oxidation and reduction reactions used by different sulphur bacteria to obtain energy.

There is a great variety of bacteria that obtain metabolic energy from reactions involving inorganic sulphur compounds, but until recently these fell into two quite distinct groups. These are (1) the sulphur-oxidizing bacteria, like thiobacilli², *Beggiatoa* and *Sulfolobus*, which carry out the



Modes of inorganic sulphur dissimilation by bacteria. *a*, Aerobic and denitrifying respiratory oxidation; *b*, anaerobic respiratory reduction; *c*, anaerobic fermentation.

aerobic respiratory (or anaerobic denitrifying) oxidation of sulphur, sulphide and other reduced-sulphur compounds to sulphate by using atmospheric oxygen (or a reduction of nitrate to nitrogen gas) as the respiratory oxidant (*a* in the figure); (2) the sulphur and sulphate-reducing bacteria, like *Desulfovibrio*, *Desulfotomaculum*, *Desulfuromonas* and the Thermoproteales³, which oxidize hydrogen or organic compounds as respiratory substrates and use sulphur compounds as the oxidants for these reactions (*b* in the figure). The metabolic and taxonomic distinction of these two processes was

blurred by the recent demonstration that *Acidianus* species can grow either aerobically as sulphur-oxidizers or anaerobically as sulphur-reducers^{4,5}, showing that these two modes of metabolism are not mutually exclusive within one organism.

The observations of Bak and Cypionka reported in this issue¹ show that these two metabolic modes can occur in one organism not only when the organism is subjected to two quite different sets of cultural conditions but can operate simultaneously in the dissimilation of inorganic sulphur substrates such as sulphite or thio-sulphate (*c* in the figure). The processes involved may thus be brought about by the simultaneous respiratory oxidation of sulphite to sulphate and reduction of sulphur or sulphite to sulphide as the respiratory electron acceptor.

Bak and Cypionka term this process fermentation, which may cause protests from some metabolic purists who will say it should properly be termed inorganic anaerobic respiration. Fermentation can be defined as an ATP-regenerating metabolic process in which degradation products of organic substrates are hydrogen donors as well as acceptors⁶, which is well illustrated, for example, by fermentation of glucose to alcohol by yeast. Mechanistically, however, a less restrictive definition can be derived, simply that the substrate molecule provides both the donor and acceptor for hydrogen transport during its metabolism, and that the oxidation states of the substrate and the sum of its products are equivalent.

Returning to the earliest definitions of fermentation by Pasteur, the process now discovered is certainly *la vie sans l'air* but, more significantly, Pasteur wrote⁷ that as the cell requires "oxygen in order to live, and that, therefore, if this element cannot be obtained in the free state, it must be taken from the fermentable material...", we can agree that the observations of Bak and Cypionka are truly the first demonstration of an inorganic fermentation. That accepted, the search for others should now commence. $\square$

1. Bak, F. & Cypionka, H. *Nature* **326**, 891–892 (1987).
2. Kelly, D.P. *Phil. Trans. R. Soc. B298*, 499–528 (1982).
3. Stetter, K.O. & Zillig, W. in *The Bacteria* Vol. 8 (eds Woese, C.R. & Wolfe, R.S.) 85–170 (Academic, Orlando 1985).
4. Kelly, D.P. *Nature* **313**, 734 (1985).
5. Segerer, A., Neuner, A., Kristjansson, J.K. & Stetter, K.O. *Int. J. Syst. Bact.* **36**, 559–564 (1986).
6. Schlegel, H.G. *General Microbiology* (Cambridge University Press, 1986).
7. Stephenson, M. *Bacterial Metabolism* 3rd edn (Longmans, London, 1949).

Donovan P. Kelly is Professor of Environmental Sciences at the University of Warwick, Coventry CV4 7AL, UK.

Relativistic astrophysics

Moving targets old and new

Virginia Trimble

MODERN astrophysics has no one holy grail to guide its searches in the way that much of particle physics is currently focused by the possibility of grand or super-unification of forces. Rather, many apparently separate goals flicker with varying intensity on the horizon. The reaching of a few of these goals, the unexpected retreat of others, and the establishment of some new ones highlighted a recent meeting*.

On the theoretical front, Robert Wald (Univ. Chicago) reported that B. F. Whiting (DAMTP, Cambridge) has, at long last, found an analytical proof of the stability of the Kerr solution against all possible perturbation modes. This means that a black hole, characterized from outside by only its mass and angular momentum, is a permanent possible final state for astronomical objects. How many actually exist?

Convincing black-hole candidates at the moment include only two or three X-ray-emitting components of binaries in Cygnus and the Large Magellanic Cloud, but Jeffrey McClintock's (Center for Astrophysics) analysis of the 1975 X-ray nova A0620-00 has added its compact component to the inventory as, in some ways, the most persuasive case. A neutron star identification for SS433, on the other hand, cannot be ruled out, according to Jonathan Katz (Washington Univ. St Louis), who also reported from a previous workshop that the evidence for a nuclear γ -ray line from SS433 has gone away. Since modelling the line (attributed to magnesium) had been so far a particularly elusive goal, its flickering out was greeted with relief.

Two effects, sufficiently ancient and honourable to have names, have been looked for yet again, one remaining undetected, the other apparently now seen in several objects. The Gunn-Peterson effect is the absorption of photons in quasar spectra shortward of 1,216 Å that should be produced by diffuse neutral hydrogen gas between us and the quasars. Charles Steidel and Wallace L. W. Sargent (California Inst. of Technology; *Astrophys. J.*, in the press) conclude that the true ultraviolet continuum of 3C9 at redshift $Z=2.016$ goes right through the top of the 'forest' of Lyman- α lines. This means that the only intervening hydrogen is that in the clouds producing those lines, the diffuse density being less than $3 \times 10^{-11} \text{ cm}^{-3}$. Diffuse neutral gas thus contributes less than one ten-thousandth of the matter density needed to close the Universe.

The Sunyaev-Zeldovich effect is also a product of diffuse gas, but in a hot, ionized state. Such gas, which reveals its presence in many rich clusters of galaxies by emitting X-rays, should Compton scatter photons from the 3 K microwave background up to much higher energies. The scattered photons will make an undetectably small contribution to visible, ultraviolet and X-ray light from the clusters, but their absence from the microwave flux in the direction of the clusters should be noticeable. Detection of the deficit would provide unambiguous evidence that the microwaves come from behind the clusters and so are truly a universal phenomenon, whereas accurate measurements of the deficit can, in principle, provide determinations of distances to the clusters and of the Hubble constant that do not depend on the usual uncertain scales. Such detailed measurements remain for the future, but Mark Birkenshaw (Harvard) reported temperature deficits of about 0.85 mK in the direction of three clusters, Abell 655, Abell 2218 and 0016+16. Previous reports of such deficits have been disputed, and so may this one.

The microwave background radiation ought to display other kinds of variations as well. Twenty years of searching have, so far, established unambiguously only a dipole anisotropy reflecting our own motion through the photon sea and none of the smaller-scale features that could probe structure in the Universe at the time galaxies were forming or earlier. The existing upper limits already severely constrain models in which galaxies evolve naturally from adiabatic perturbations in baryon density. David Wilkinson's (Princeton Univ.) review presented two tentative detections of fluctuations in the 3 K background, though neither pertains to the galaxy formation scale (near 10^4) and each has a possible alternative interpretation.

First, R. D. Davies and Anthony Lasenby (Jodrell Bank; also Davies, R.D. *et al. Nature* **326**, 462-465; 1987) have seen 0.1 mK differences between sky regions about 8° apart (the scale corresponding to the size of parts of the Universe in communication with themselves at the time the photons were last scattered). The possibility that the fluctuations merely represent lumpiness in bremsstrahlung emission from our own Galaxy (synchrotron can already be ruled out by comparison with existing maps) can and will be tested by mapping at 5 GHz as well as the original 10 GHz.

Also, Edward Fomalont and colleagues (NRAO) have looked on the 10-30''

scales mappable with the Very Large Array telescope and find fluctuations as large as a 1 mK, some of which undoubtedly simply result from faint sources scattered over the sky. Computer simulations, based on extrapolations of counts of radio sources at other wavelengths and brighter fluxes, indicate that there is a real, residual lumpiness in the background radiation. Again, maps at additional frequencies can help to separate the effects.

Among the receding goals are a resolution of the solar neutrino problem and a persuasive model of galaxy formation. Two leading contenders for reducing the detectable electron neutrino flux at earth are neutrino oscillations (which convert up to two-thirds of the emitted flux to undetectable μ or τ neutrinos in the solar mantle or *en route*) and various mechanisms that cool the solar core so that fewer neutrinos are produced in the first place. Because the latter reduces the high-energy neutrino fraction more efficiently than does the former, an experiment with a lower energy threshold than the 0.814 MeV of Raymond Davis's chlorine detector ought to be able to distinguish between them. Unfortunately, calculations of the sensitivity of the most promising candidate experiment (a gallium detector), as presented by Davis (Univ. Pennsylvania) show that some forms of each explanation predict essentially the same total count rates in gallium. Though the counts would come from different proportions of high- and low-energy neutrinos in the two cases, there is no way to distinguish them.

Attempts to combine assorted physical processes to produce galaxies without ruffling up the 3 K background too much have multiplied beyond counting (let alone reading!) in recent years. At the Texas Symposium in Jerusalem two years ago, the combination of hot dark matter to provide most of the gravitating mass and cosmic strings to trigger condensation looked very promising, but, two years downstream, computer modelling by Adrian Melott (Univ. Kansas) and others strongly suggests structures made this way will not much resemble the ones we actually see.

In two cases, theory and observation have advanced to meet each other so expeditiously that it is hard to say which was the goal and which the pursuit. It has been clear for many years that the radio-emitting regions of giant elliptical galaxies were, on average, smaller in the past, but just how much and why has been debated. G. Krishna and Paul Wiita (Georgia State Univ.) have modelled the interactions among intergalactic, galactic and relativistic jet gases to predict that radio source diameter should decrease with redshift ($= \text{distance} = \text{look-back time}$) as $(1+Z)^{-1}$. And P. Katgert (Leiden Observatory) presented a survey of radio galaxies at varying redshifts whose sizes do scale as $(1+Z)^{-1}$.

*13th Texas Symposium on Relativistic Astrophysics, Chicago, Illinois, 15-19 December 1986.

A second rapid convergence has occurred for binary X-ray systems in globular clusters. The first well-established binary period has been with us for less than a year (11.3 min for the source in NGC6624; W. Priedhorsky *et al.* *IAU Circ.* No. 4247; 1986). Jonathan Grindlay (Center for Astrophysics) announced the second (8.5 h for the source in M15, based on archival data from HEAO-1 and Einstein) and, in the same talk, provided a model in which the two represent different evolutionary phases of the same sort of object. The longer orbit period is identified as the earlier evolutionary phases, in which rapid

mass transfer is driven by nuclear evolution of the companion star. The establishment and expulsion of a common envelope around the two stars drains angular momentum from the system, leaving a short-period object in which mass transfer is driven much more slowly by gravitational radiation. Thus we expect that the M15 source should eventually resemble the NGC6624 one! □

Virginia Trimble is Professor of Physics, University of California, Irvine, California 92717 and Visiting Professor of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

Natural selection

Did Komodo dragons evolve to eat pygmy elephants?

Jared M. Diamond

WHEN Komodo dragons, the largest living lizards, were discovered on a small, remote Indonesian island in 1910, there was astonishment that these carnivores (see figure), up to 3 m long and 100 kg or more in weight, could have escaped attention for so long. How did such a large carnivorous lizard obtain its food, and why had it evolved on Komodo? Accumulating information suggests that oras (to use the local name for these lizards) evolved to prey on now-extinct pygmy elephants.

Oras belong to the genus *Varanus* (monitor lizards), which arose in Indonesia and then spread to produce several dozen modern species from Australia to Africa¹. The name Komodo dragon (*V. komodoensis*) for oras is somewhat misleading, as the animals inhabit Komodo, the adjacent, much larger island of Flores and three other nearby small islands, all of which formed one land-mass in the Pleistocene, when the sea level was low. The most detailed information about their diet comes from Auffenberg's field work¹. Although hatchling oras eat insects and lizards, medium-sized animals eat rodents and birds and full-sized adults prey on and scavenge large mammals — mainly deer and pigs, but also buffalo, horses, goats

and, occasionally, people. Victims are either caught asleep or ambushed from about 1 m away. In several cases oras have been observed to seek out pregnant ungulates close to term, in order to seize the newborn, precipitate a miscarriage or attack the incapacitated mother.

A clue to the evolution of oras on the Flores island group is the former existence in Australia of an even larger monitor lizard, *Megalania prisca*, which was up to 6 m long and 2,000 kg in weight^{2,3}. Although Pleistocene Australia supported many large marsupial herbivores (giant kangaroos, wombats and rhinoceros-like diprotodonts), only two species of large marsupial carnivores, the Tasmanian wolf (*Thylacinus*) and the marsupial lion (*Thylacoleo*), evolved there to prey on those herbivores³. The large carnivore niche left otherwise vacant by Australia's inaccessibility to the placental carnivores of Asia was filled by the evolution of *Megalania*, which disappeared, together with marsupial lions and large marsupial herbivores, at the end of the Pleistocene.

Between the Australian and Asian continental shelves lies the island region known to zoogeographers as Wallacea, which consists of the Flores group, Timor, Celebes and other islands. Like Australia,

the islands of Wallacea were inaccessible to Asia's placental carnivores, but Wallacea was also inaccessible to Australia's marsupial carnivores. But monitor lizards, being good swimmers, were able to reach Flores and evolved into the ora to fill the large carnivore niche there, just as *Megalania* did in Australia.

This reconstruction faces two serious objections. First, all the large mammal species on which oras now prey were introduced to Flores by humans after BC 3000 (refs 4,5). What did oras eat before then? During the Pleistocene, Flores supported two species of now-extinct elephants, the small *Stegodon trigonocephalus florensis* and the pygmy *S. sompoensis* (also called *S. timoriensis*)^{6,7}. Hence Auffenberg's suggestion¹ that oras evolved as a specialized predator on pygmy elephants, formerly the sole large prey available on Flores. Because the spread of humans from Asia through Wallacea reached Australia and New Guinea by BC 4000 (ref. 8), humans must have reached Flores at least by then. But it remains unknown how much earlier humans were on Flores, how long it took them to exterminate the pygmy elephants and what oras were eating during the time (if any) between the elephants' demise and the arrival of domestic ungulates around BC 3000.

Second, pygmy stegodonts formerly occupied at least three other Wallacean islands (Timor, Sumba and Celebes) plus the Philippine island of Mindanao, as well as Flores^{6,7}. Why are oras confined to the Flores group? In fact, there is an unsubstantiated modern report of oras from Sumbawa (the next large Wallacean island west of Flores), and large monitor lizards variously attributed to *V. komodoensis* or related species are found as fossils in Timor and Java¹. Thus, ora-like monitors were formerly much more widespread in the Indo-Australian region. Populations other than those on the Flores group must have been eliminated by climatic changes, extermination of their native prey species, and depredations by humans and their associated dogs, pigs and monkeys. Had Alfred Russel Wallace been able to make his famous journey through the Malay Archipelago 50,000 years ago, he might have found giant monitors harassing pregnant pygmy elephants on many islands. □

1. Auffenberg, W. *The Behavioural Ecology of the Komodo Monitor* (Univ. Florida Press, Gainesville, 1981).

2. Rich, T. & Hall, B. in *Vertebrate Zoogeography and Evolution in Australasia* (eds Archer M. & Clayton G.) 391-394 (Hesperian, Australia, 1984).

3. Martin, P.S. & Kline, R.G. *Quaternary Extinctions* (Univ. Arizona Press, Tucson, 1984).

4. Glover, I.C. *World Archaeol.* **9**, 42-61 (1977).

5. Bellwood, P. *Prehistory of the Indo-Malaysian Archipelago* (Academic, New York, 1985).

6. Hooijer, D.A. *Mod. Quat. Res. SE Asia* **1**, 37-46 (1975).

7. Sartono, S. *Mod. Quat. Res. SE Asia* **5**, 57-63 (1979).

8. Groube, L. *et al.* *Nature* **324**, 453-455 (1986).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The monstrous Komodo dragon, with a hearty appetite.

Second, does Fig. 1 really show "saturation" of GTP γ S binding? The amount of binding shown at "saturation" is 60 pmol (after 10 min incubation of 100 nM radio active ligand with 50 μ g membrane protein); the paper does not specify whether this means 60 pmol per tube or 60 pmol per mg protein. The first is impossible, because each tube contained only 5 pmol radiolabelled ligand in the indicated volume of 50 μ l. The second is inconsistent with Fig. 2, because the indicated binding affinity ($K_d = 166$ nM), number of binding sites (1.8 nmol mg $^{-1}$) and ligand concentration used (100 nM) would specify occupancy — at equilibrium — of 677 pmol binding sites per mg protein; at the indicated concentration of membrane protein (1 mg ml $^{-1}$), this would be a 6.8-fold higher concentration of occupied sites than the total concentration of ligand present. Something is wrong. One guess is

that equilibrium was never reached in the experiment of Fig. 1. The apparent saturation of GTP binding sites probably reflects binding of almost all the available ligand, because radiolabelled ligand concentration, rather than concentration or affinity of binding sites, was limiting.

Third, is it likely that the GTP γ S binding sites described in Fig. 2 mediate the IL-2-stimulated GTPase activity reported in Table 1 and Fig. 3? Let us assume (as seems to be the case for other GTP-binding proteins) that the affinity of the protein for binding GTP γ S ($K_d = 166$ nM) indicates a K_m of similar magnitude for hydrolysing GTP. If so, the 4 pmol GTP hydrolysed per 100 μ g protein per min at 100 nM substrate (in the presence of IL-2; Table 1, Fig. 3) is consistent with a V_{max} of 106 pmol GTP hydrolysed per mg protein per min. At 1.8 nmol GTP binding sites per mg protein, the average GTPase molecule, even when stimulated, would hydrolyse no more than one molecule of GTP every 17 min.

These simple calculations indicate serious flaws in the evidence presented by Evans *et al.* Lymphocyte membranes seem to contain impossibly large amounts of an extraordinarily sluggish GTPase; internal inconsistencies suggest errors in the performance or interpretation of the binding studies. These flaws reduce the case for a GTP-binding protein as mediator of IL-2 action to two findings: IL-2 stimulates membrane GTPase, and pertussis toxin treatment blocks this stimulation. The latter result strongly predicts that the toxin will block biological effects of the IL-2 on intact T lymphocytes. Correction of the binding analysis and tests of the prediction are clearly required.

HENRY R. BOURNE

Department of Pharmacology,
University of California,
San Francisco,
California 94143, USA

Once again, how many reactor accidents?

SIR—No doubt, after Harrisburg and Chernobyl, the probability of another serious nuclear reactor accident in the next ten years is disturbingly high, but how high is it really? The estimation of Islam and Lindgren¹, who assumed a Poisson process with mean failure rate r , suffers from an incorrect prior probability, 47 years after Jeffreys² demonstrated that only dr/r works properly for rate parameters, and 18 years after Jaynes³ gave a simple group-theoretical proof. Edwards⁴ recognizes this but disregards the correct prior, too. He takes the likelihood function of the Poisson model, $P(K|r, T) = e^{-rT} (rT)^K / K!$, that is the conditional probability of K accidents, given the average rate r and the total past operation

Table 1 Probabilities of k accidents in the next $t = 3,740$ reactor years, given $K = 2$ accidents in $T = 4,000$ reactor years

k	$P(k t, K, T)^*$	$P(k t, K, T, \tau)^{\dagger}$
0	0.267	0.492
1	0.258	0.311
2	0.187	0.131
3	0.121	0.046
≥ 4	0.167	0.020

* Without numerical prior information.

$\dagger$ With $1/\tau = 10^{-4}$ accidents per reactor year estimated *a priori*.

time T , and substitutes $r = -\ln P(0|r, t)/t$, where t is some future operation time. From a plot of the logarithm of the resulting expression (the 'support') for $K = 2$, $T = 4,000$ reactor years and $t = 3,740$ reactor years (the numbers in ref. 1, for 374 reactors operating at present) he concludes that "all we can really say is that the probability of an accident in the next ten years is unlikely to be less than a quarter", claiming that this result is superior to what physicists can offer. Schwartz⁵ gives a more definite frequentist result, but again without arriving at the required probability $P(0|t, K, T)$ of no accident, given t , K and T . This probability is easily derived as follows.

The posterior distribution of r is proportional to the product of likelihood function $P(K|r, T)$ and prior dr/r . Normalizing to unity, one gets the probability that the rate parameter lies in an infinitesimal interval dr at r ,

$$p(r|K, T) dr = \Gamma(K)^{-1} e^{-rT} (rT)^{K-1} dr/r, \quad 0 < r < \infty.$$

The expectation value, $\langle r \rangle = K/T$, agrees perfectly with common sense, and the relative uncertainty, $\delta r / \langle r \rangle \equiv \sqrt{(\text{var } r)} / \langle r \rangle = 1/\sqrt{K}$, has the familiar dependence on the number of events. The joint probability that both " r in dr " and " k within t " are true is $p(r|K, T) dr$ times $e^{-rt} (rt)^k / k!$, whence the probability of k future accidents, given t , K and T but regardless of r , is found by integration over all r :

$$P(k|t, K, T) = \frac{(K+k-1)!}{k!(K-1)!} \frac{T^K t^k}{(T+t)^{K+k}},$$

$$k = 0, 1, 2, \dots$$

This is the distribution we want. The expectation value $\langle k \rangle = Kt/T = \langle r \rangle t$ makes sense again, the relative uncertainty is $\delta k / \langle k \rangle \equiv \sqrt{(\text{var } k)} / \langle k \rangle = \sqrt{\{(1+T/t)/K\}}$. With the numbers of ref. 1 one obtains the second column shown in the table, and $\langle k \rangle = 1.87$, $\delta k / \langle k \rangle = 1.02$. It is hardly surprising that after two serious accidents within the past 4,000 reactor years we must expect about two more serious accidents in the next 3,740 reactor years. This expectation is quite uncertain, however. There is still a chance that no accident at all will occur, but the odds against this are about 3 to 1. On the

other hand the probability of three or more accidents is 29 per cent.

This result, based only on past experience and on stationarity of the Poisson process, reflects neither numerical information about failure rates which may be available *a priori*, nor possible improvements in safety technology and reactor operations. If one has an *a priori* estimate $1/\tau$ for the failure rate, but no information about its uncertainty, one can invoke the maximum entropy principle³ according to which the prior $e^{-r\tau} \tau dr$ is the most objective choice possible. Its mean is just $1/\tau$ but otherwise it resembles dr/r , decreasing monotonically as r increases. Using this more informative prior instead of dr/r one finds

$$P(k|t, K, T, \tau) = \frac{(K+k)!}{K! k!} \frac{(T+\tau)^{K+1} t^k}{(T+\tau+t)^{K+k+1}},$$

$$k = 0, 1, 2, \dots$$

with mean $\langle k \rangle = (K+1)t/(T+\tau) = \langle r \rangle t$ and relative uncertainty $\delta k / \langle k \rangle = \sqrt{\{(1+(T+\tau)/t)/(K+1)\}}$. With $1/\tau = 10^{-4}$ core melt accidents per reactor year (the number quoted in ref. 1 from the German Risk Study⁶) and K , T and t as before, one gets the third column of the table and $\langle k \rangle = 0.80$, $\delta k / \langle k \rangle = 1.26$, that is an expected number of accidents slightly smaller than one, but again with large uncertainty. The chance of no serious accident in the next ten years is then about fifty-fifty.

One could refine⁷ the maximum entropy prior by also using the uncertainty information about τ given in ref. 6, but we shall not do so. The main conclusion of ref. 1 remains valid in any case: failure rate estimates of 10^{-6} accidents per reactor year are much too optimistic in view of past experience.

This is a physicist's solution of the problem considered in ref. 1, valid for very small samples ($K \geq 1$ without prior information, $K \geq 0$ with the more informative maximum entropy prior). It employs nothing but the basic sum and product rules of probability theory (and, in the first step, Bayes' theorem, their immediate consequence) and modern techniques for the assignment of prior probabilities which have a solid foundation in either group theory or information theory.

F. H. FRÖHNER

Institut für Neutronenphysik und
Reaktortechnik,
Kernforschungszentrum Karlsruhe,
Postfach 3640,
D-7500 Karlsruhe, FRG

- Islam, S. & Lindgren, K. *Nature* **322**, 691–692 (1986).
- Jeffreys, H. *Theory of Probability* (Clarendon, Oxford, 1939).
- Jaynes, E. T. *IEEE Trans. Systems Sci. Cybern.* **SSC-4**, 227–241 (1968); *Papers on Probability, Statistics and Statistical Physics* 114–130 (Reidel, Dordrecht, 1983).
- Edwards, A. W. F. *Nature* **324**, 417–418 (1986).
- Schwartz, J. *Nature* **324**, 622 (1986).
- Deutsche Risikostudie (BMFT/TÜV Rheinland, Bonn, 1979).
- Fröhner, F. H. *Risk Analysis* **5**, 217–225 (1985).

Migration of mid-ocean-ridge volcanic segments

Hans Schouten*, Henry J. B. Dick* & Kim D. Klitgord†

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

† US Geological Survey Woods Hole, Massachusetts 02543, USA

The propagation of small-offset volcanic spreading-centre segments along mid-ocean ridge crests may reflect absolute motion of the plate boundary relative to the underlying mesospheric frame. Such a relationship could be caused by a purely vertical flow of the mantle under spreading centres and would have value in constraining past plate motions from non-transform trends generated during along-ridge propagation and in linking the major-element variability of oceanic crust and upper mantle to the bulk composition and temperatures of mantle ascending under mid-ocean ridges.

MID-OCEAN ridges are offset by transform zones ranging in size from less than a kilometre to several hundred kilometres. Large-offset transform zones usually remain stationary relative to the plate boundary for many tens of millions of years, and divide the ridges into a series of relatively straight sections with lengths of tens to hundreds of kilometres. The sections are themselves segmented systematically by smaller offsets on a length scale of 30–80 km, which is related to spreading rate¹. These small-scale volcanic segments have their origin in the asthenosphere under ridges where, at some depth beneath the axis, the mantle melt (magma) segregates from the partially molten mantle (the sub-axial asthenosphere), concentrates at evenly spaced locations and drains quickly through cracks or diapirically to crustal magma chambers and associated volcanic spreading-centre segments^{1–3}.

Topographic lineaments

The volcanic segments have a pronounced topographic signature which is preserved in older crust as parallel topographic lineaments. Although these traces of past spreading-centre segmentation are sometimes developed parallel to flow-lines of relative plate motion⁴, indicating that the volcanic segments did not propagate along the plate boundary, in other places these traces are oblique to flow-lines, forming broad V's about the spreading axis, indicating that the volcanic segments did propagate^{5,6} (see Fig. 1a and b). These V-shaped non-transform traces on the segmentation scale should not be confused with the topographic and magnetic lineaments that mark propagating rifts^{7,8}. These occur on a much larger length scale (typically hundreds of kilometres) and the non-transform pseudo-fault traces often indicate rift propagation in opposing directions. Such a geometry suggests a localized driving mechanism (such as the Galapagos and Cobb hotspots in the vicinity of the Galapagos and Juan de Fuca groups of propagating rifts).

The multiple non-transform traces associated with the volcanic segments, on the other hand, form a consistent set of features indicating propagation of a string of mid-ocean-ridge volcanic segments in one direction along the ridge crest. Such a geometry suggests a more regional driving mechanism (such as a sub-axial flow of the asthenosphere beneath the ridge⁵). In order to distinguish between these two classes of propagation phenomena we shall refer to the latter as migration.

In the slowly spreading North Atlantic, distant from known hotspots, sets of multiple non-transform lineaments have been mapped by detailed ridge-crest surveys at two places. At these locations, the chains of topographic highs and lows can be traced directly to the present-day segmentation of the ridge (Fig. 1a and b). The consistency of the non-transform trends indicates that in both cases the segments have been migrating along the spreading axis as a group. The non-transform traces of both Atlantic areas in Fig. 1 are symmetrical about an axis orthogonal to the spreading direction, but not about the oblique ridge crest.

Calculated migration velocities that match these non-transform trends then depend on whether migration has occurred in a direction orthogonal to the spreading direction with symmetric spreading, or parallel to the oblique ridge crest with asymmetric spreading. Depending on the assumptions made, the migration velocity of the volcanic segments along the ridge is calculated at 49° N as 2.7 or 3.1 km Myr⁻¹ respectively and at 26° N as 3.5 or 3.7 km Myr⁻¹.

On the fast-spreading East Pacific Rise (EPR), recent Seabeam investigations have revealed a fine-scale volcanic segmentation of the ridge axis^{9–11}. The rather sparse off-axis survey data available show near-axis volcanic ridges or seamount chains with a significant non-transform direction^{11,12} (Fig. 1c and d). Recent petrological studies of zero-age basalts dredged from the EPR axis suggest that mid-ocean-ridge volcanic segments are migrating southwards at overlapping spreading centres and saddle points near 12.9° N, 9.1° N, 8.6° N and 5.5° N (refs 13, 14). Magnetic studies of overlapping spreading centres near 11.7° N and 9.1° N also suggest a southward migration^{15,16}. The petrological and magnetic conclusions are consistent with the direction of migration indicated by the off-axis non-transform trends. This section of the EPR is distant from any known hotspots. The Pacific non-transform trends in Fig. 1 indicate southward migration at similar velocities though separated by more than 500 km.

The seamount chains in Fig. 1c and d appear to occur only on one side of the spreading axis, and are thus different from the symmetrical parallel traces in the Atlantic. Furthermore, the near-axis chains are generally shallower than the rise axis, suggesting voluminous off-axis volcanic activity. Lonsdale¹¹ attributed the chain near 5.5° N to southward migration of the dextral overlapping spreading centre at 5.5° N. The chain at 10° N was related by Crane³ to an axial high at 9.83° N, but might equally be associated with a small dextral ridge offset nearby (see Fig. 1c). The one-sided distribution of the two Pacific non-transform traces might thus be related to southward migration of dextral offsets between volcanic segments and analogous to the one-sided failed rifts in the propagating rift geometry¹⁷ (see also Fig. 2a). The shallow depth of these near-axis seamount chains then suggests that young failed rift zones would be a locus of significant off-axis volcanic activity.

Flow of asthenosphere

If the origin of such observed non-transform traces lies in a sub-axial asthenosphere flow⁵, it is now necessary to ask whether that flow has any regularity. The orientation of the non-transform traces of Fig. 1 relates qualitatively to the absolute plate motions deduced from hotspot traces, shown schematically in Fig. 2. This concept was tested using the absolute plate motion model AM1–2 of Minster and Jordan¹⁸. We find that a southward sub-axial asthenosphere flow of 28 km Myr⁻¹, estimated from the Pacific non-transform traces of Fig. 1, correlates with the

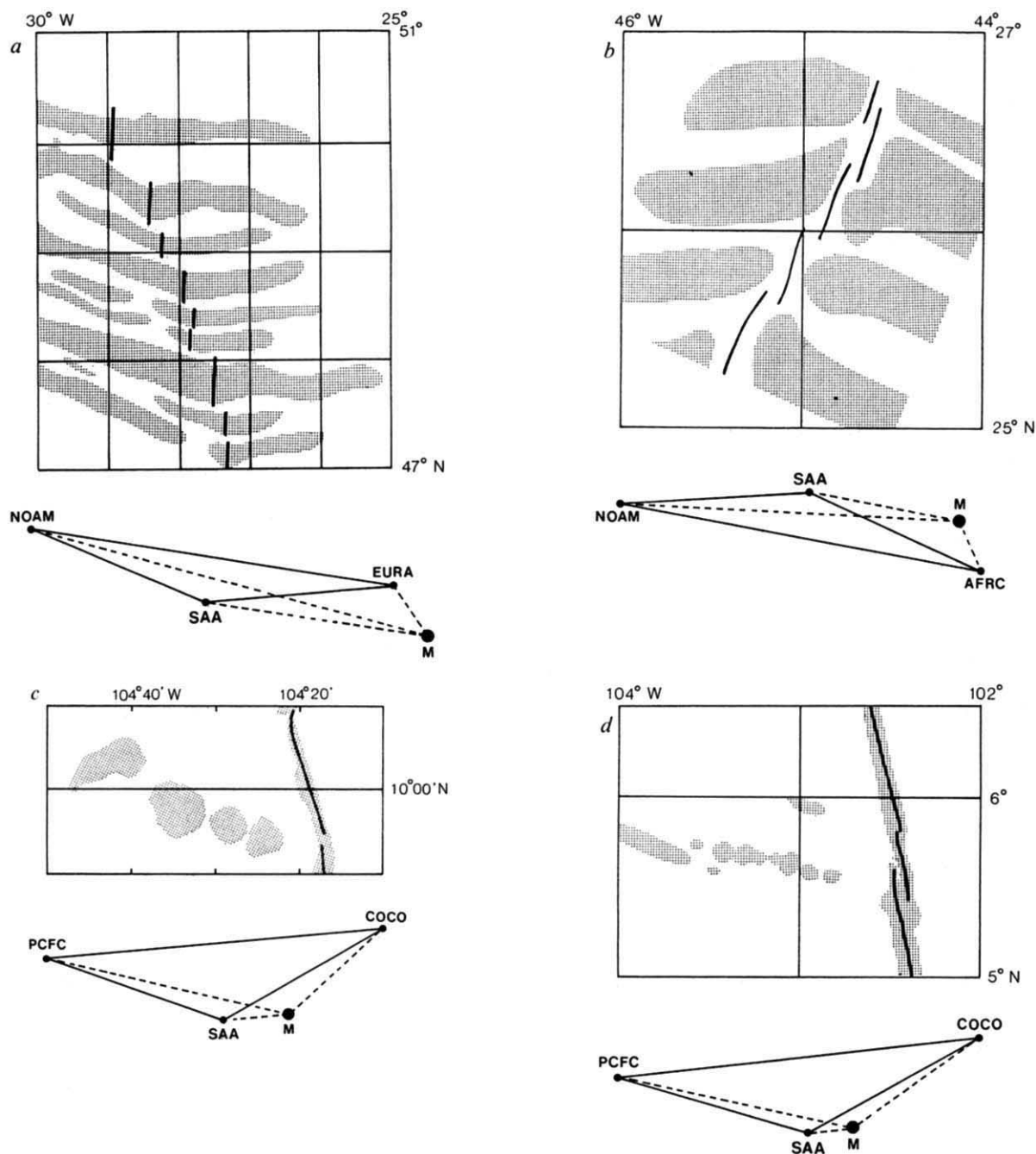


Fig. 1 Non-transform patterns of ridge crest indicating migration of volcanic segments in the Atlantic and eastern Pacific. *a*, North Atlantic plate boundary (NOAM-EURA). Stippled areas show interpreted trends of topographic belts which clearly form broad V's about the spreading axis (after ref. 5). The belts can be traced directly to the present-day segmentation of the active ridge (short heavy lines, modified after ref. 5). The oblique angles of the recent (<5 Myr) trends indicate southward migration velocities of $\sim 3 \text{ km Myr}^{-1}$. *b*, Central Atlantic plate boundary (NOAM-AFRC). Stippled areas show recent (<5 Myr) topographic highs separated by pronounced topographic lows which can be traced to small offsets between the active spreading centre segments (modified after ref. 6). The angles of these trends indicate northward migration velocities of $\sim 3.5 \text{ km Myr}^{-1}$. *c*, East Pacific plate boundary near 10°N (PCFC-COCO). Stippled areas show water depths shallower than 2,600 m, the heavy lines show the spreading-centre segments (after ref. 12). The near-axis seamount chain can be explained alternatively as the one-sided trace of an axial high at 9.83°N migrating southward³ or off-axis volcanic activity marking a young failed rift zone caused by southward migration at $\sim 28 \text{ km Myr}^{-1}$ of the small dextral offset between segments at 9.9°N . *d*, East Pacific plate boundary near 5.5°N (PCFC-COCO). Stippled areas show water depths shallower than 3,000 m and the heavy lines the active spreading centre segments (after ref. 11). The major non-transform seamount chain on the young Pacific plate can be related to the dextral overlapping spreading centre at 5.5°N and its migration southward at $\sim 25 \text{ km Myr}^{-1}$ (ref. 11). A minor topographic lineament to the north of this chain can similarly be related to a small dextral offset between segments at 5.85°N .

The non-transform trends in (*c*) and (*d*) are separated by more than 500 km but yield similar migration velocities in the same direction suggesting a regional sub-axial flow of asthenosphere beneath the boundary between the Pacific and the Cocos plate. The velocity-space diagrams (explained in Fig. 2) show the non-transform trends (SAA to plate) predicted for the four ridge locations from Pacific absolute plate motion vectors¹⁸ (*c* and *d*) and from adjusted Atlantic vectors (*a* and *b*) which are consistent with the observed non-transform trends and with the absolute plate motion model AM1-2 within its limits of error as shown in Table 1. This analysis indicates that a sub-axial asthenosphere flow, which would cause volcanic segments to migrate, could be driven by the absolute motion of the plates.

Table 1 Atlantic absolute motion vectors

Plate	θ ° N	σ_θ deg.	ϕ ° E	σ_ϕ deg.	ω deg. Myr ⁻¹	σ_ω deg. Myr ⁻¹
AFRC*	18.76	33.93	-21.76	42.20	0.139	0.055
EUR*	0.70	124.35	-23.19	146.67	0.038	0.057
NOAM*	-58.31	16.21	-40.67	39.62	0.247	0.080
AFRC†	30.00	—	-34.00	—	0.191	—
EUR†	33.78	—	-49.80	—	0.092	—
NOAM†	-42.86	—	-48.64	—	0.243	—

* Absolute motion model AM1-2 of Minster and Jordan¹⁸.

† Data from this study. These vectors obey the relative motion model RM2 of Minster and Jordan¹⁸ and are used to construct the velocity space diagrams for Fig. 2a and b.

ridge-parallel component of apparent motion of the mesosphere relative to a symmetrically spreading East Pacific Rise as shown by the velocity diagrams for the two Pacific locations in Fig. 1.

The AM1-2 is not well constrained for the North Atlantic plates. The deduced migration rates of 3 km Myr⁻¹ southward at 49° N and 3.5 km Myr⁻¹ northwards at 26° N are, however, consistent with the Atlantic absolute motion vectors within their limits of error. We have used the non-transform trends, the relative motion model RM2 (ref. 18) and the assumption of symmetric spreading to recalculate present-day absolute motion vectors of the North American, Eurasian and African plates, which are given in Table 1, together with the AM1-2 vectors and their estimated uncertainties. These adjusted vectors were used to construct the velocity diagrams for the two Atlantic locations in Fig. 1.

From this analysis we propose that the sub-axial asthenosphere flow, which causes volcanic segments to migrate along ridge axes, is driven by the absolute motion of the plates. The migration equals the axis-parallel component of motion of the hot-spot frame relative to the spreading plates at their boundary (see Fig. 2). We now examine the consequences of this postulate.

Upwelling of mantle

First, a correlation between the migration of volcanic segments along the ridge and absolute plate motions implies that the volcanic segments are closely linked to the mesosphere and that, similar to the mesosphere, the framework from within which they originate must have horizontal absolute motions much slower than the motions of the plates. If the volcanic segments represent persistent melting anomalies within the upper mantle under ridges as argued by Whitehead *et al.*², then this part of the upper mantle must be below the zone of shearing that, in a broad sense, must separate the mesosphere and its motion from those of the lithospheric plates. Since the depth of segregation of melt from the shallow mantle may be no more than a few tens of kilometres below ocean spreading centres (see ref. 2), this in turn implies that, directly under ridges, the mesospheric frame exerts a strong influence to within a short distance of the surface of the Earth. On this basis, the mantle which rises toward mid-ocean ridges would rise essentially vertically in the mesospheric framework to a very shallow level and release its melt before it enters the shear zone between mantle and the lithospheric plates beneath the ridge. The absolute motion vector of the sub-axial asthenosphere (M-SAA in Fig. 2c) then represents the progression through the ascending mantle of the locus of highest melt production or, the locus of fastest and shallowest mantle upwelling between the separating plates.

Next, consider a kinematic model (shown in the ridge-fixed frame in Fig. 3) in which the mantle rises vertically under mid-ocean ridges that move laterally over the mantle. Decoupling between vertical upwelling of mantle and horizontal plate motions is assumed for simplicity to take place in an infinitesimally thin shear zone at the base of the lithosphere.

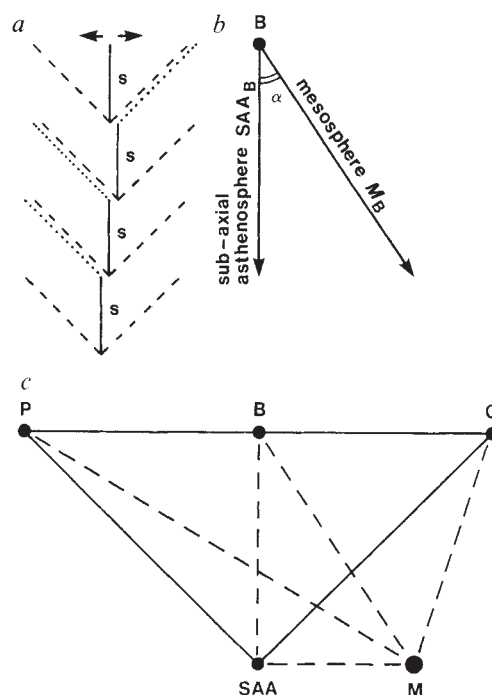


Fig. 2 *a*, Small-offset volcanic spreading-centre segments (*s*) propagate at a uniform velocity in the same direction along the boundary between two separating lithospheric plates, leaving multiple and parallel non-transform traces in older crust away from this boundary. The dashed lines represent the pseudo-faults⁷, the symmetric traces of transition zones between consecutive volcanic segments where topographic lows can be expected. The one-sided dotted lines represent the failed rift zones¹⁷, the deactivated tail ends of the active spreading-centre segments left on one plate or the other by propagation of the next volcanic segment. The young failed rift zones could be a locus of off-axis volcanic activity forming one-sided non-transform volcanic ridges or seamount chains. The angle of the non-transform trends is a measure of the propagation velocity. *b*, The propagation of a string of small-offset volcanic spreading-centre segments (here referred to as migration) can be explained by a flow of the sub-axial asthenosphere beneath the ridge which is directly related to the motion of the plate boundary in the fixed-hotspot mesospheric reference frame. Motions of the sub-axial asthenosphere (SAA_B) and the mesospheric frame (M_B) are referenced to the plate boundary (*B*). $SAA_B = M_B \cos \alpha$. *c*, The geometry in velocity space of the postulated motion of the sub-axial asthenosphere (SAA) relative to the two separating plates (*P* and *Q*), the plate boundary (*B*, assuming symmetric seafloor spreading), and the fixed-hotspot mesospheric reference frame (*M*). The heavy line connecting *P* and *Q* denotes the relative plate motion, the heavy lines connecting SAA with *P* and *Q* denote the postulated non-transform trends. The dashed lines denote the 'sub-axial asthenosphere flow' (*B* to SAA), and the absolute motions of the various reference frames.

Conservation of mass requires that the vertical flow of mantle (V_z) in the wake of the separating and growing lithospheric plates is then proportional to the slope of the base of the lithosphere (θ) and to the normal-to-axis velocity (V_p) of each plate relative to the 'fixed' mantle or, $V_z = V_p \tan \theta$ (ref. 19). The fastest upwelling of shallow mantle and consequent highest melt production by decompression occurs beneath the ridge and thus moves relative to the mesosphere with the normal-to-axis component of absolute plate boundary motion (see also Fig. 2).

The ratio between total spreading rate and normal-to-axis velocity of the spreading axis over the mantle (which we define as the mantle consumption ratio) is then a measure of the total mantle upwelling required to balance the growth of the spreading lithospheric plates. This ratio scales the flow field under both spreading plates and scales the depth to which spreading

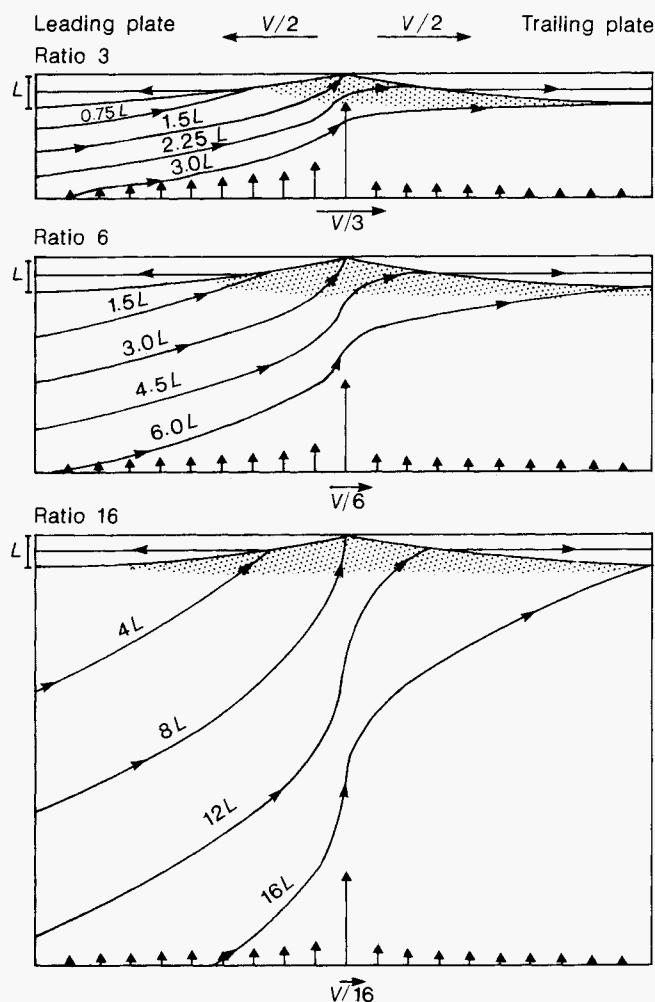


Fig. 3 Kinematic model of whole-mantle upwelling under symmetrically spreading ridges moving laterally over the mantle (shown in the ridge-fixed frame). This model assumes purely vertical flow of mantle material (in the absolute motion reference frame) and a decoupling between vertical mantle upwelling and horizontal plate motion at the base of the lithospheric plates. Vertical mantle flow (vertical arrows) is then proportional to the slope of the base of the lithosphere and the normal-to-axis velocity of each plate relative to the 'fixed' mantle (after ref. 19). Horizontal arrows beneath each diagram give the apparent horizontal motion of the mantle reference frame relative to the ridge. The stream lines, which show how the underlying mantle would appear to move to an observer on the ridge, predict significant asymmetries in paths and melt histories of mantle material that underplates the leading or the trailing lithospheric plate, and a path and melt history of mantle material that forms the upper mantle under oceanic crust that is dependent on the mantle consumption ratio (total spreading rate divided by the normal-to-axis velocity of a symmetrically spreading ridge relative to the mantle). Initial depths of mantle material now upwelling to balance the growth of the spreading lithospheric plates, are given in thickness of older lithosphere (L , assumed to be constant outside the diagrams). Diagrams for mantle consumption ratios 3 and 16 predict the shallower upwelling under the American-Antarctic Ridge and South-west Indian Ridge, and deeper upwelling under the southern Mid-Atlantic Ridge near the Bouvet Triple Junction, respectively. Stippled areas give schematically the melting region for a mantle of uniform composition, determined by the particular geotherm chosen as the initial boundary condition at the start of adiabatic mantle upwelling on the left of the diagrams. The lowest degree of mantle melting occurs near the base of the melting region (that is, the mantle solidus), the highest degree occurs directly beneath the thin lithosphere at the spreading axis and is related to initial depth or, mantle consumption ratio.

plates must ideally 'consume' underlying mantle (see Fig. 3). For higher consumption ratios, deeper mantle will be added to the lithosphere; for a ratio of x , the spreading plates would ultimately have consumed upper mantle equal to roughly x thicknesses of older lithosphere (L in Fig. 3 which is assumed to be constant outside the diagrams). In the case of a thicker shear zone at the base of older lithosphere, mass balance will require a greater mantle upwelling than with the infinitesimal shear zone in Fig. 3. Typically, the present-day ratios calculated from the AM1-2 model vary from ~ 2 to ~ 7 or higher, and they may vary considerably along any boundary.

Degree of mantle melting

The degree of partial melting of the mantle emplaced to the base of the crust at spreading centres is largely determined by its total depth of upwelling. For a mantle of homogeneous bulk composition and no contaminating fluids leaking into the shallow mantle from below, the degree of melting of the shallowest mantle under spreading centres then becomes a function of its initial depth and the initial geotherm (that is, the depth and geotherm before significant mantle upwelling due to an approaching mid-ocean-ridge system). Because this initial geothermal gradient was probably more conductive than adiabatic in the thermal boundary layer, in the shallower mantle under the plates, mantle ascending adiabatically from a shallower initial depth will be less molten than mantle ascending from greater initial depths (see ref. 20).

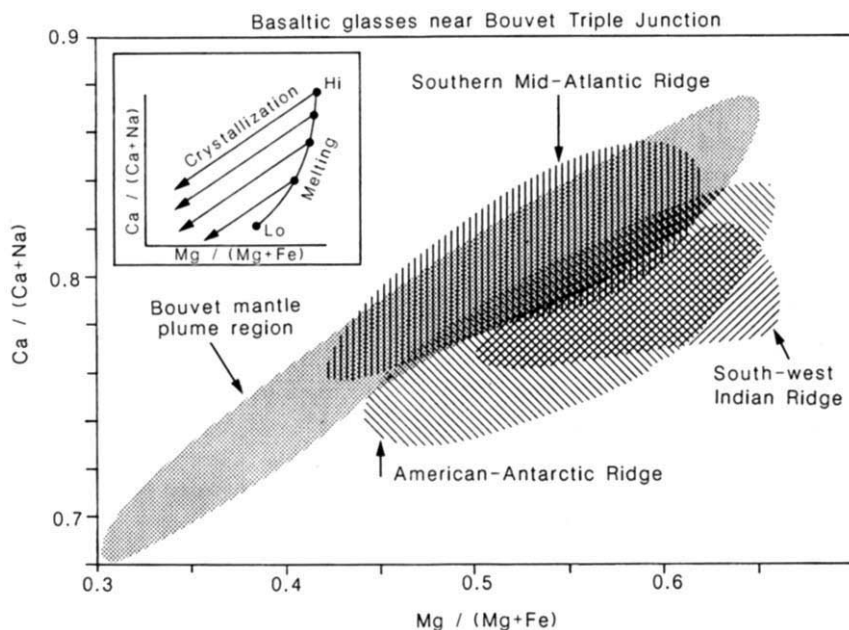
Some of the deepest mantle upwelling (and highest degree of mantle melting) is predicted for the southern Mid-Atlantic Ridge (SMAR) between South America and Africa near the Bouvet Triple Junction where the ridge has apparently remained nearly fixed in the absolute motion frame since the Mesozoic break-up of the continent of southern Gondwana (see ref. 21). In contrast, shallower upwelling (and lower degree of mantle melting) is predicted for the American-Antarctic Ridge (AAR) and the south-west Indian Ridge (SWIR) immediately to the west and east of the triple junction where the trailing plate (Antarctica) has remained relatively stationary. (A stationary ridge implies a very high mantle consumption ratio, a stationary trailing plate implies a ratio of 2.) The AM1-2 model yields present-day mantle consumption ratios of 16.4 for the SMAR (for the equatorial MAR this ratio is 3.5) and 2.7 for both AAR and SWIR near the triple junction.

Basaltic lavas dredged along these ridges exhibit systematic differences in major-element composition (Fig. 4) related to differences in degree of melting of the shallow mantle under the ridges with which the melt equilibrates before segregation and eruption²². Figure 4 shows the fields for the $\text{Ca}/(\text{Ca} + \text{Na})$ and $\text{Mg}/(\text{Mg} + \text{Fe})$ ratios of 574 basaltic glasses from these ridges, with glasses from the SWIR separated into those dredged far from and near to the Bouvet mantle plume. As can be seen, the glasses from the Bouvet mantle plume region, and those from the southern end of the MAR plot are along approximately the same linear trend across the diagram. In contrast the glasses from the SWIR, far from the Bouvet mantle plume, and those from the AAR plot are along a similar parallel linear trend lower in the diagram.

This plot is particularly useful in that it uses simple binary oxide ratios and effectively separates liquidus trends for different primary magmas. In addition, the effects of fractional crystallization are easily distinguished from those due to varying degrees of melting (or initial source composition). As shown in the inset, variable degrees of melting of the mantle produces a relatively small variation in $\text{Mg}/(\text{Mg} + \text{Fe})$ of the partial melt, but a large variation in $\text{Ca}/(\text{Ca} + \text{Na})$. In contrast, fractional crystallization of olivine and plagioclase from a melt produces large variations in both ratios in a sense opposite to that for increasing degrees of mantle melting.

The systematic offset in Fig. 4 between the fields for the AAR and SWIR, and the field for the SMAR reflect the similar low

Fig. 4 Composition of basaltic glasses dredged near the Bouvet Triple Junction of the South American, African, and Antarctic plates. As shown by the inset cartoon, increasing degree of mantle melting produces melts with slowly increasing $Mg/(Mg+Fe)$ and rapidly increasing $Ca/(Ca+Na)$. Subsequent fractionation of melts representing different degrees of melting during crystallization and cooling, will produce systematically offset trends of decreasing $Mg/(Mg+Fe)$ and $Ca/(Ca+Na)$. The basalts dredged from the American-Antarctic Ridge (2° – 18° W) and South-west Indian Ridge away from the Bouvet mantle plume (5° – 25° E), both plot along lower trends in the diagram, reflecting lower degrees of melting of the parent mantle, than basalts dredged from the southern Mid-Atlantic Ridge (51° – 55° S). The basalts from the southern Mid-Atlantic Ridge plot along virtually the same trend as basalts from the Bouvet mantle plume region (1° – 5° E), reflecting a high degree of melting and deep mantle upwelling for both regions. The isotopic composition of basalts from these two regions is significantly different, precluding an origin for the southern Mid-Atlantic Ridge basalts related to the Bouvet plume (A. Le Roex, personal communication). The plume-like deep upwelling suggested by the high degrees of melting under the southern Mid-Atlantic Ridge and the shallower upwelling predicted from lower degrees of melting of the mantle under the other two ridges are explained with the geometry of absolute plate motions particular to this area and the existence of a significant thermal boundary layer beneath the lithosphere of southern Gondwana before the Mesozoic break-up of that continent.



degrees of melting of the mantle beneath the AAR and SWIR, and the higher degree of melting beneath the SMAR, as was predicted from our model. Furthermore, the high $Ca/(Ca+Na)$ ratios of basalts dredged from the Bouvet mantle plume region and from the SMAR are indistinguishable, confirming the plume-like deep mantle upwelling and consequent highest degrees of mantle melting beneath the SMAR that this kinematic model (Fig. 3) would predict. The same high $Ca/(Ca+Na)$ ratios against $Mg/(Mg+Fe)$ also characterize the zero-age basalts dredged from the EPR (J. Natland, personal communication) suggesting that their mantle parent also rose from great initial depths. This is consistent with the low estimates of net absolute motion of this plate boundary over the past 100 Myr (refs 23, 24) for which our model independently predicts deep mantle upwelling in the East Pacific region.

Markers of absolute plate motion

A final consequence of our postulate is that non-transform traces produced by migration of volcanic ridge segments provide estimates of absolute plate motion, present and past, independent of the hotspot traces. Such estimates can be important in reconstructing motions of plates with ill-defined hotspot traces.

Considering the affinity to the mesospheric frame of the migration of volcanic segments along Pacific and Atlantic mid-ocean ridges, a predominantly vertical mantle flow beneath these ridges seems likely. This is supported by the correlation between the degree of melting of the shallow mantle under ridges and net absolute plate boundary motion. If our observations and interpretations of migration of mid-ocean ridge volcanic segments can be applied to present and past ridges on a global scale, it could provide a new and compelling argument for whole-mantle convection (see ref. 25). A comprehensive examination of the major-element variability of zero-age basalts could then provide important constraints on the bulk composition and temperatures of mantle upwelling under mid-ocean ridges.

Many of the ideas presented here have been nurtured by the weekly WHOI Geodynamics seminars sponsored by the Center for Analysis of Marine Systems (CAMS) of the Woods Hole Oceanographic Institution with funds from the Exxon Education Foundation. We thank D. Epp, R. C. Kerr, J. Phipps Morgan, J. Natland, W. Normark, A. Trehu, J. A. Whitehead and especially J. R. Cann for constructive criticism. This work was supported by the Office of Naval Research and the NSF. Pam Foster and Maureen Carragher processed the manuscript.

Received 29 December 1986; accepted 11 February 1987.

- Schouten, H., Klitgord, K. D. & Whitehead, J. A. *Nature* **317**, 225–229 (1985).
- Whitehead, J. A., Dick, H. J. B. & Schouten, H. *Nature* **312**, 146–148 (1984).
- Crane, K. *Earth planet. Sci. Lett.* **72**, 405–414 (1985).
- Schouten, H. & Klitgord, K. D. *Earth planet. Sci. Lett.* **59**, 255–266 (1982).
- Johnson, G. L. & Vogt, P. R. *Geol. Soc. Am. Bull.* **84**, 3443–3462 (1973).
- Rona, P. A., Harbison, R. N., Bassinger, B. G., Scott, R. B. & Nalwalk, A. J. *Geol. Soc. Am. Bull.* **87**, 661–674 (1976).
- Hey, R. & Vogt, P. *Tectonophysics* **37**, 41–52 (1977).
- Hey, R. N. & Wilson, D. S. *Earth planet. Sci. Lett.* **58**, 167–188 (1982).
- Macdonald, K. C. & Fox, P. J. *Nature* **302**, 55–58 (1983).
- Macdonald, K. C., Sempere, J.-C. & Fox, P. J. *J. geophys. Res.* **89**, 6,049–6,069 (1984).
- Lonsdale, P. *Geol. Soc. Am. Bull.* **96**, 313–327 (1985).

- Fornari, D. J., Ryan, W. B. F. & Fox, P. J. *J. geophys. Res.* **89**, 11069–11083 (1984).
- Hekinian, R. *et al. Mar. geophys. Res.* **7**, 359–377 (1985).
- Langmuir, C. H., Bender, J. F. & Batiza, R. *Nature* **322**, 422–429 (1986).
- Sempere, J.-C., Macdonald, K. C. & Miller, S. P. *Geophys. J. R. astr. Soc.* **79**, 799–811 (1984).
- Sempere, J.-C. & Macdonald, K. C. *Eos* **67**, 1227 (1986).
- Hey, R., Duennebie, K. & Morgan, W. J. *J. geophys. Res.* **85**, 3647–3658 (1980).
- Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331–5354 (1978).
- Davis, E. E. & Karsten, J. L. *Earth planet. Sci. Lett.* **79**, 385–396 (1986).
- McKenzie, D. *J. geophys. Res.* **25**, 713–765 (1984).
- Morgan, W. J. *Tectonophysics* **94**, 123–139 (1983).
- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. *Earth planet. Sci. Lett.* **69**, 88–106 (1984).
- Jurdy, D. M. & Gordon, R. G. *J. geophys. Res.* **89**, 9927–9936 (1984).
- Schult, F. R. & Gordon, R. G. *J. geophys. Res.* **89**, 1,789–1,800 (1984).
- Loper, D. E. *J. geophys. Res.* **90**, 1809–1836 (1985).

A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands

Xavier Estivill, Martin Farrall, Peter J. Scambler, Gillian M. Bell, Kim M. F. Hawley, Nicholas J. Lench, Gillian P. Bates, Helena C. Kruyer, Patience A. Frederick, Philip Stanier, Eila K. Watson, Robert Williamson* & Brandon J. Wainwright

Cystic Fibrosis Genetics Research Group, Department of Biochemistry, St Mary's Hospital Medical School, University of London, London W2 1PG, UK

A genomic sequence close to the cystic fibrosis locus with the characteristics of an HTF island has been selectively cloned and characterized. Two markers flanking this sequence, which is conserved throughout mammalian evolution, show a very much greater disequilibrium than that found with any existing marker. A single mutational event accounts for most cases of cystic fibrosis. The sequence is expressed, and is a candidate for the cystic fibrosis gene.

MOLECULAR biology offers powerful approaches to the analysis of inherited diseases where the biochemical defect is unknown. For chronic granulomatous disease, Duchenne muscular dystrophy and retinoblastoma the existence of chromosomal deletions and/or translocations focused attention on particular chromosome regions, and candidate sequences for the gene altered in the disease have been identified¹⁻³. However, there are many inherited diseases such as cystic fibrosis (CF) with no associated visible changes in chromosome structure, where localization by genetic linkage using affected families gives a map resolution only to within several million base pairs (bp). In such cases it is necessary to combine several different molecular and genetic approaches to avoid extensive DNA sequencing and stepwise chromosome walking.

Cystic fibrosis is a common severe autosomal recessive gene defect affecting the Caucasian population, with a carrier frequency of 1 in 20 (ref. 4); in spite of intensive research the basic defect causing CF remains unknown. Genetic linkage studies between CF, the protein marker *PON*⁵, and the DNA markers *met*⁶, *J3.11* (*D7S8*)⁷, *917* (*D7S15*)⁸, *TCRB*⁷, *COL1A2*⁹, *7C22* (*D7S16*)¹⁰ and *B79a* (*D7S13*)^{11,12} have localized the CF mutation to chromosome 7q22-31. From data from a collaborative study the recombination fraction between *J3.11* and CF was estimated to be 0.003, and between *met* and CF to be 0.004, suggesting that *J3.11* and *met* flank CF¹³. The study shows that the proportion of unlinked (heterogeneous) families is very small; >95% of families show linkage to 7q markers. A further linkage study with data from CEPH (Centre d'Étude du Polymorphisme Humain) reference pedigrees and CF families¹⁴ gave strong support to the order *TCRB-J3.11-CF-met-7C22-B79a-COL1A2*.

Using chromosome-mediated gene transfer (CMGT) we have constructed several mouse/human hybrid cell lines that contain transgenomes including the markers most closely linked to CF¹⁵. These cell lines have proved useful for the exclusion of candidate genes for the CF mutation^{16,17}, and for the physical location of new markers for this region¹⁸. To analyse the human component of these cell lines we have constructed several cosmid libraries^{19,20}. Many of the cosmid clones have been analysed and ordered into contiguous overlapping maps ('contigs')^{19,21}. The contigs in turn have been studied using pulsed-field gel electrophoresis (PFGE)²² and rare-cutter restriction mapping²⁰. Two cosmid clones which detect polymorphic loci in strong linkage disequilibrium with CF have been isolated¹⁹.

Many vertebrate genes are preceded by a region of DNA rich in the non-methylated dinucleotide CpG, and containing clusters of sites for CpG-methylation sensitive restriction enzymes²³. These sequences, known as HTF islands (*HpaII* tiny

fragments), have a sequence length of between 500 and 2,000 bp, and often include the first exons as well as upstream sequences 5' (and sometimes 3') to the associated coding gene^{23,24}. HTF islands can be used as one set of signals marking coding sequences. We describe here the successful use of a 'rare-cutter cosmid library' in the identification of a genomic region, with the characteristics of an HTF island, in high level of linkage disequilibrium with CF. The associated conserved sequence is a strong candidate for the CF gene.

Putative HTF island in CF region

To identify regions of the genome which contain coding sequences, we have used a selection technique to isolate clones derived from non-methylated CpG-rich regions of the transgenomes. The procedure involves the construction of a cosmid library using a rare-cutter restriction enzyme sensitive to methylation (non-methylated CpG rare-cutter cosmid library).

DNA from the cell line CIId was digested with *XmaIII* and *HindIII* and the library constructed using the method outlined in Fig. 1a. CIId is a subclone of CII which has lost some human chromosome 7 sequences, including *J3.11*, and contains about 1 megabase (Mb) of human DNA; CII contains ~4 Mb of human DNA^{15,19}. The *XmaIII* (*EagI*) restriction enzyme was chosen because it cuts relatively infrequently and is methylation-sensitive²⁵. Therefore the construction of a cosmid library from a transgenome 1 Mb in length, using *XmaIII* as cloning site, allows the selection both of a set of spaced 'milestones' at intervals of several hundred kilobases (kb) along this region of chromosome 7, and of putative coding sequences by the identification of HTF islands. Recent studies have shown that methylation-free islands are maintained in transgenic mice²⁶, suggesting that they should remain unmethylated in the cell lines obtained by CMGT.

The cosmid library constructed contains 10⁵ recombinants. Five different human cosmid clones, designated cNX.1 to cNX.5, were identified using two rounds of screening by hybridization with labelled placental human DNA as the probe. Three of the five clones were assigned to chromosome 7 using a panel that contains DNA from cell line clone 21 (CI 21) (which contains the entire human chromosome seven in a mouse background²⁷), clone 36IS (CI 36) (with 7p1.3-7qter as its only human component²⁸), CII DNA, and total human and mouse DNAs.

A 7 kb *HindIII* single-copy fragment (designated pXV-2a) from cosmid cNX.2 hybridizes to a 7-kb fragment in *HindIII* digests of total human DNA and DNA from CI 21, CI 36 and CII (Fig. 1b). This demonstrates that pXV-2a is located at human chromosome 7q22-7q31 and that the CIId cell line does not contain gross rearrangements or deletions at this site.

* To whom correspondence should be addressed.

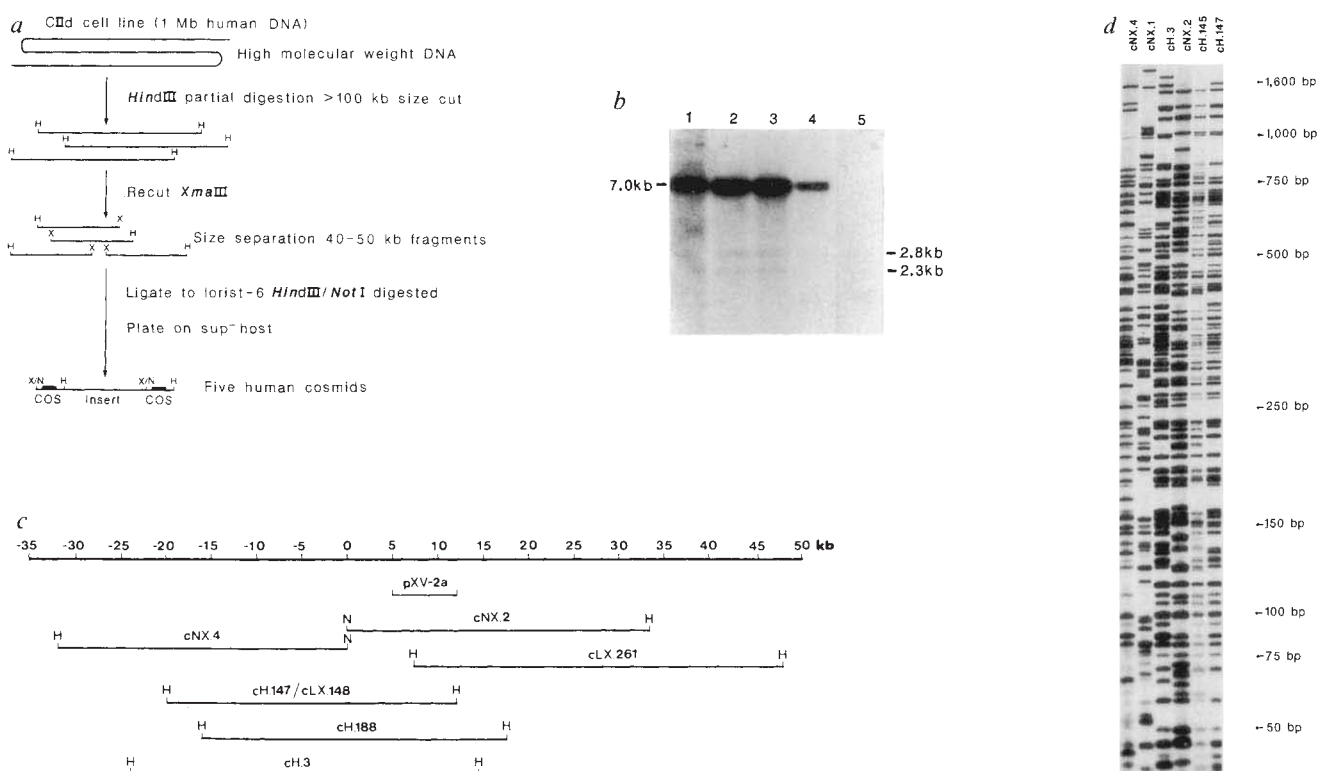


Fig. 1 Construction and analysis of the *Xma*III/*Hind*III cosmid library. **a**, Strategy for the preparation of the *Xma*III/*Hind*III cosmid library. High-molecular-weight DNA was partially digested with *Hind*III and recut with *Xma*III. The 40–45-kb fraction was isolated and ligated to the *Not*I/*Hind*III sites of Lorist-6 (from Dr P. Little), a modification of Lorist-B³⁹ and packaged *in vitro*⁴⁰, using Gigapack (Stratagene). Five human cosmids were identified by hybridization with labelled placental human DNA as the probe. DNA from each cosmid was isolated and digested with *Xma*III/*Hind*III and *Not*I/*Hind*III, and analysed as described^{19,20}. The average insert size was 40 kb. Single-copy DNA fragments were subcloned into the plasmid vector pUC13 and into the cosmid vector Lorist-6. Subclones of single-copy fragments were then used individually as hybridization probes. H, *Hind*III; X, *Xma*III. **b**, Southern blot hybridization of the probe pXV-2a, a 7-kb *Hind*III fragment from cosmid cNX.2, to *Hind*III digests of 5 µg of genomic and hybrid-cell-line DNAs. Lane 1, human genomic DNA; lane 2, C121; lane 3, C1 36; lane 4, C11d; lane 5, mouse genomic DNA. *Hind*III-digested λ DNA was used for size markers. Hybridization to a fragment of 7 kb was observed for samples 1 to 4. Hybridization to fragments of 2.8 and 2.3 kb were detected for the samples 2 to 5. **c**, Schematic representation of the cNX.4-cNX.2 contig. Probe pXV-2a was used to screen a cosmid bank of 550 cosmids^{19,20}. Four contiguous cosmids were identified in one step and two more cosmids were detected which had previously been isolated and characterized from EIII (refs 19 and 20). H, *Hind*III; N, *Not*I. Coordinates indicate the distance in kilobases (kb) starting at zero for the position of the *Not*I site. **d**, Polyacrylamide gel electrophoresis analysis of 5 cosmids from the cNX.4-cNX.2 contig. Cosmid cNX.1 was used as negative control for fingerprinting.

Methods. In **a**, transduction was carried out using *Escherichia coli* ED8767 as the host. Bacteria were plated on nylon membranes on kanamycin-containing L-broth agar at ~5,000 colonies per 140 mm diameter plate. Duplicate filters were hybridized in 6×SSC, 5×Denhardt's, 0.5% SDS and 50 µg ml⁻¹ heparin for 24 h at 65 °C. Sonicated and denatured human placental DNA labelled using random oligonucleotide primers with [α -³²P]dCTP (ref. 41) was used as probe. Membranes were washed in 0.1×SSC and 0.1% SDS, and autoradiographed at -70 °C for 24 h with an intensifying screen against preflashed film. General methods for Southern blot analysis are described elsewhere⁴². The labelled probe pXV-2a was competed with total human DNA as described¹⁹, hybridized to membranes and washed as a stringency of 0.1×SSC, 0.1% SDS. Autoradiography conditions were as described above for cosmid screening. Chromosome walking techniques were as described^{19,20}. Methods used for cosmid fingerprinting were a modification of those of Coulson *et al.*²¹. Cosmid miniprep DNA (5 µl) was mixed with a reaction mixture, containing 2×*Hinf*I reaction buffer, 20 mM spermidine and 2 units of *Hinf*I, and incubated at 37 °C for 1 hour. Then 5 µl of a reaction mixture containing 3 µCi of [α -³⁵S]dATP and 10 units of cloned Moloney murine leukaemia virus reverse transcriptase, were added and the reaction was incubated for a further hour at 37 °C. Labelled *Hinf*I-treated pBR322 fragments were used as size markers. Reactions were loaded onto a 6% acrylamide/bis-acrylamide gel in Tris-borate buffer and run for 150 min at 500 V. Autoradiography was as described previously²¹.

Subsequently, pXV-2a was used to screen a cosmid bank of human clones from the cell line EIII (refs. 19 and 20) (EIII contains ~10 Mb of human DNA, largely from chromosome 7 (ref. 15). Eight overlapping cosmids were detected (Fig. 1c). A second clone from the *Xma*III/*Hind*III cosmid library, cNX.4, was found to overlap with cosmids from the same contig, but not with cNX.2. The cosmid contig including the supposedly 'linked' *Xma*III cosmids cNX.4-cNX.2 extends over a region of about 80 kb (Fig. 1c). We have fingerprinted these cosmids using a modification of the approach of Coulson *et al.*²¹ and have obtained both the overlap and map order (Fig. 1d). The isolated contig does not overlap with another contig 160 kb long that we have characterized¹⁹ which includes *met*, giving a

minimum distance of 100 kb between pXV-2a and *met*.

Cosmids cNX.2 and cNX.4 each contain a *Not*I site, which is consistent with the cloning procedure as the *Xma*III site includes the internal six base pairs of the 8-bp *Not*I recognition sequence. Cosmid cHX.148, which contains both a *Not*I and a *Bss*HII site (detected using *Not*I and *Bss*HII oligonucleotides²⁰) overlaps with both *Xma*III cosmids (Fig. 1c) and confirms that cNX.2 and cNX.4 each contain one half of a single *Not*I site and are adjacent to one another.

A rare-cutter restriction map of cosmids cHX.148/cHX.147, containing the *Not*I site, was constructed using both oligonucleotide probes and restriction digests for several rare-cutters (Fig. 2a and b). A fragment of 1,500 bp contains 11 restriction

Table 1 Distribution of alleles for the D7S8, *met*, XV-2c and CS.7 loci

Locus and alleles	CF		Alleles		Relative risk	χ^2_1	Δ
	1	2	1	2			
pJ3.11 <i>MspI</i> Alleles: 1 = 4.2 kb 2 = 1.8 kb	73	61	58	75	1.50	3.2	0.11
<i>pmetD</i> <i>TaqI</i> Alleles: 1 = 6.0 kb 2 = 4.4 kb	90	24	89	23	0.97	0.01	-0.006
<i>pmetH</i> <i>TaqI</i> Alleles: 1 = 7.5 kb 2 = 4.0 kb	98	49	94	45	0.96	0.03	-0.01
pXV-2c <i>TaqI</i> Alleles: 1 = 2.1 kb 2 = 1.4 kb	195	18	93	109	12.4	103.0	0.49
pCS.7 <i>HhaI</i> Alleles: 1 = 0.67 kb 2 = 0.47 kb	1	71	38	28	0.015	54.0	-0.62

Δ , Standardized disequilibrium coefficient. The pXV-2c data include 17 Finnish and 10 Danish families (54 CF chromosomes: A1 51; A2 3, and 54 normal chromosomes: A1 19; A2 35), overall relative risk = 26.8; $\chi^2_1 = 42.0$; $\Delta = 0.62$.

sites for 8 such enzymes. This cluster of sites, and the sensitivity of the genomic DNA to *Xma*III (as shown in the construction of the cosmid library), suggests that the non-methylated CpG dinucleotide is present at a high frequency in this region of the genome, one of the criteria for an HTF island.

To define more precisely a non-methylated clustered target area for the enzymes *Hpa*II and *Hha*I (which are methylation-sensitive), genomic DNA from several individuals was digested with *Sac*I, *Sac*I/*Hpa*II and *Sac*I/*Msp*I. The DNA was hybridized with a 2.0-kb *Not*I/*Hind*III fragment, derived from cosmid cNX.4, containing the putative 'island' (designated pCS.7). Fragments smaller than 300 bp were obtained for *Hpa*II and *Msp*I digests (Fig. 2c). A similar result was obtained using *Hha*I instead of *Hpa*II (data not shown). The pattern obtained using pXV-2a against the same blot showed only one *Hpa*II non-methylated site, in a total length of more than 20 kb of genomic DNA to which this 7-kb fragment hybridizes (data not shown).

To map the location of the putative HTF island, and particularly its relation to the *met* locus, we used two subclones of the cNX.4-cNX.2 contig (pCS.7 and pXV-2c), which flank the HTF *Not*I site, to probe Southern blots of *Not*I-digested human DNA separated by PFGE (Fig. 3a). Probe *pmetD* recognizes a simple two-allele restriction fragment length polymorphism (RFLP) with *Not*I on PFGE, with bands at 820 and 470 kb (Fig. 3b). In patients with CF the less-frequent allele was found in 2 of 10 individuals (allele frequency 10%).

Plasmid pXV-2c is polymorphic for the same *Not*I site as *met*, and hybridizes to the same 820-kb fragment, but in this case the less-frequent allele generates a fragment of 350 kb rather than 470 kb (Fig. 3c). The polymorphic *Not*I site is therefore in the 820-kb *Not*I fragment and is flanked by the probes *pmetD* and pXV-2c. The *Not*I site within the cNX.4-cNX.2 contig, of which pXV-2c is a part, must be situated at one end of this 820-kb fragment. The *met* locus has been reported as being within 150 kb of the *Not*I site at the other end of the 820-kb fragment²⁹; therefore XV-2c and *metD* must be separated by at least 650 kb. XV-2c is about 15 kb away from the putative HTF island. When pCS.7 (which maps at the other half of the *Not*I site of the cNX.4-cNX.2 contig) was hybridized to the same blot, a *Not*I fragment larger than 1 Mb was observed (not

shown). From these data we can infer that we have jumped 650–750 kb from the *met* locus.

Sequence in disequilibrium with CF

Two probes, pXV-2c and pCS.7, derived from the cosmid contig, detect frequent RFLPs. To assess the genetic and physical proximity of XV-2c and CS.7 to CF, these markers were used for segregation analysis in both our reference CF families and in families ascertained by demonstration of recombination between CF and the most closely linked markers, and for studies of linkage disequilibrium.

Probe pXV-2c detects an RFLP (A1 0.47, A2 0.53) in *Taq*I digests (Fig. 4a) and pCS.7 detects an RFLP (A1 0.56, A2 0.44) in *Hha*I digests (Fig. 4b). No recombination with CF was observed with either pXV-2c or with pCS.7 in six informative families showing a crossover between CF and J3.11, or 7C22 and/or B79a. Two *met* recombinant families were not informative in the relevant parent for either pCS.7 or pXV-2c.

Combining the crossover data presented here with the data for probe-probe genetic order¹⁴, the PFGE result given above, and the data on the *met* activation in the transgenic cell lines^{15,30}, the order of loci is: 7qter-J3.11-(CF/CS.7/XV-2c)-*met*-7C22-B79a-7cen.

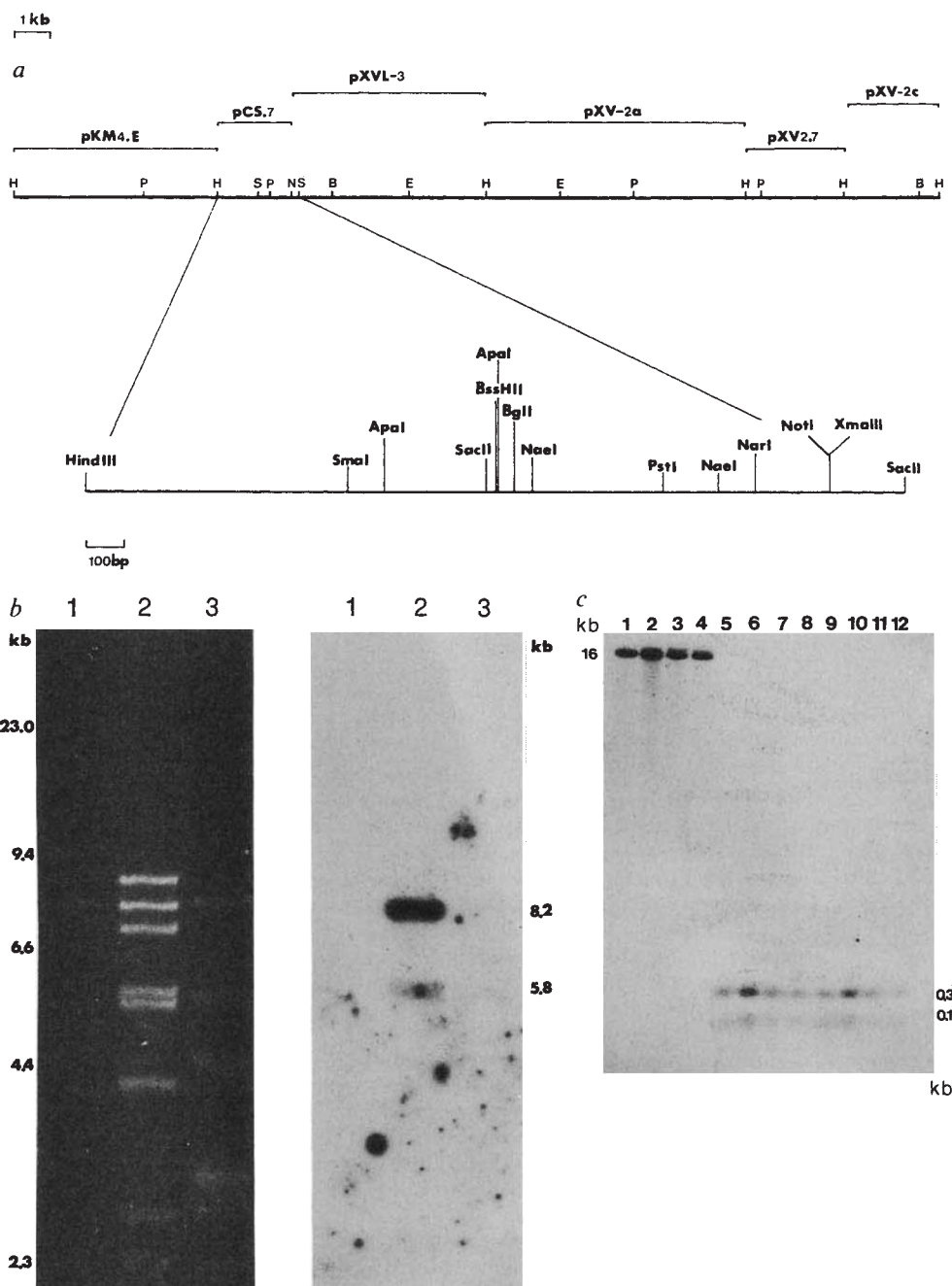
In 195 out of 213 chromosomes studied the CF mutation is associated with the pXV-2c A1 allele (2.1 kb), and in 70 out of 71 chromosomes with the pCS.7 A2 allele (0.47 kb) (Table 1). This suggests that CF is closer to CS.7 than to XV-2c.

Table 2 XV-2c/CS.7 Haplotype distribution of the CF alleles

Haplotype	pXV-2c <i>TaqI</i>	pCS.7 <i>HhaI</i>	Normal alleles	(%)	CF alleles	(%)
A	1	1	9	(14)	0	(0)
B	1	2	22	(34)	66	(94)
C	2	1	28	(43)	0	(0)
D	2	2	6	(9)	4	(6)
Total			65	(100)	70	(100)

Alleles are defined in Table 1.

Fig. 2 Characterization and mapping of the putative HTF island. *a*, Schematic restriction map of part of the cNX.4-cNX.2 contig and rare cutting restriction map of the HTF island. Points of cleavage for the restriction enzymes *Hind*III (H), *Pst*I (P), *Sac*II (S), *Not*I (N), *Bam*HI (B) and *Eco*RI (E) are indicated. Cloned DNA segments from human recombinant cosmids are indicated on the top. DNA was isolated and restriction mapped as described⁴². *b*, Detection of the restriction sites *Not*I and *Bss*HII using oligonucleotides as probes²⁰. To identify human cosmids containing rare-cutter restriction sites, minipreped cosmid DNAs from a cosmid bank^{19,20} were digested with *Hind*III or *Hind*III/*Sac*I, the fragments analysed by agarose gel electrophoresis and transferred to nylon membranes (left). *Not*I (eight bases long) and *Bss*HII (seven bases long) oligonucleotides were used as probes. Hybridization conditions for *Not*I (right) have been described²⁰. For *Bss*HII (data not shown) hybridization and washing were carried out at 24 °C. Lane 1, cosmid LX.147; lane 2, cLX.148; lane 3, cLX.149. Sizes of vector band (5.8 kb) and of human fragment (8.2 kb) are indicated. Size markers were fragments of *Hind*III-digested λ DNA. *c*, Hybridization of subclone pCS.7 to Southern transfer of DNA from four individuals digested with *Sac*I (lanes 1-4), *Sac*I/*Hpa*II (lanes 5-8) and *Sac*I/*Msp*I (lanes 9-12). Methods were as in Fig. 1. The 1-kb ladder was used as size marker (not shown).



The only CF family studied in which the CF chromosome is associated with pCS.7 allele A1 is also the only 'crossover family' (KN) in which XV-2c, and *met* and 7C22, recombine with CF. This family also shows a second apparent recombination event at the CF locus in a previous generation. We do not know whether sampling or laboratory error, multiple recombinations, or heterogeneity explains this exceptional family.

A weak but significant non-random allelic association (linkage disequilibrium) had previously been demonstrated for both J3.11/CF and *met*/CF haplotypes¹³. However, the degree of disequilibrium varied between different laboratories, and our own data showed linkage equilibrium in our population for both *met* and J3.11 (refs 13 and 31). To assess the strength of association for all closely linked markers in our families, disequilibrium between CF and marker alleles was evaluated separately for pXV-2c, pCS.7, *pmet*H, *pmet*D and pJ3.11 using contingency table analyses; the standardized disequilibrium coefficient (Δ) was calculated according to Hill and Robertson³². Significant linkage disequilibrium with CF was found only for

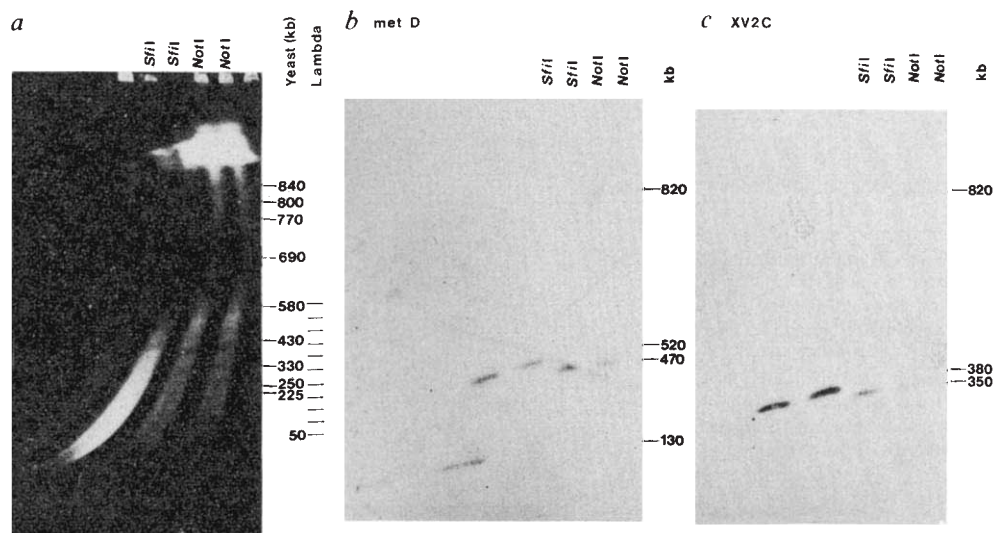
the allelic systems defined by pXV-2c ($\Delta = 0.49$; $\chi^2_1 = 103$) and pCS.7 ($\Delta = -0.62$; $\chi^2_1 = 54$) (Table 1). If the families from Denmark and Finland are analysed separately for XV-2c/CF (see foot of Table 1), the strength of association for this group is $\Delta = 0.62$; $\chi^2_1 = 42$, which may reflect the fact that these Scandinavian families are from a more homogenous population than the other families studied.

Linkage disequilibrium between the marker loci for the haplotypes *met*H/XV-2c, *met*D/XV-2c, J3.11/XV-2c, and CS.7/XV-2c was estimated. No disequilibrium was found for J3.11/XV-2c, *met*H/XV-2c or *met*D/XV-2c on either normal or CF chromosomes. The marker loci CS.7 and XV-2c are in strong linkage disequilibrium both for normal and CF chromosomes (statistics calculated for pooled chromosomes: $\Delta = -0.62$; $\chi^2_1 = 57$) (Table 2). As a recombinant between CF and XV-2c has occurred in the KN family, chromosome phase may not be unambiguously assigned and no haplotypes are included for the KN family in Table 2.

Most (94%) CF chromosomes are of haplotype B. Because

Fig. 3 Hybridization of *pmetD* and *pXV-2c* to *NotI*- and *SfiI*-digested genomic DNA. *a*, The ethidium bromide stained pulse field gel of genomic DNA, cut with *NotI* or *SfiI*. Genomic DNA was from human lymphocytes and after digestion was fractionated using orthogonal field agarose gel electrophoresis. Marker lanes are yeast chromosomes and concatemerized wild-type λ DNA molecules. *b*, Hybridization with *pmetD*. In this individual *pmetD* is heterozygous for the *NotI* polymorphism and recognizes fragments of 820 kb and 470 kb. *c*, Hybridization with *pXV-2c*, which is heterozygous for the same *NotI* polymorphic site and recognizes the same 820 kb fragment and one of 350 kb.

Methods. Human lymphocytes were isolated from venous blood by spinning over Lymphoprep (Nyegaard), washed twice in phosphate-buffered saline and suspended in 0.5% low melt agarose blocks (Seaplaque FMC) at a concentration of 5×10^5 cells per block. The cells were lysed, and DNA purified, by incubating the blocks in 0.25 M EDTA pH 8.0, 2 mg ml⁻¹ Proteinase K and 1% (w/v) *N*-lauroylsarkosine at 50 °C for 48 h. These procedures for block preparation, storage and restriction enzyme digestion have been described⁴³. Digests with *NotI* and *SfiI* (New England Biolabs) were performed according to the manufacturer's recommendations. Yeast chromosomes were prepared as previously described⁴⁴. DNA of λ was embedded in 0.5% low melting point agarose blocks at a concentration of 8 μ g per block in water, and allowed to concatemerize at 37 °C for 3.5 h in 3 $\times$ SSC. DNA was fractionated on an LKB Pulsaphor with a pulse time of 65 s at 330 V for 40 h at 13.5 °C and transferred to a nylon membrane and hybridized as described⁴⁵. Probes were labelled by random oligonucleotide priming to a specific activity in excess of 10^8 d.p.m. μ g⁻¹ DNA⁴¹. Filters were washed to a stringency of 0.1 $\times$ SSC and autoradiography was for 1–4 days at –70 °C as in Fig. 1.



the majority of RFLPs are common to more than one population, and probably arose before the racial divergence of man, the most plausible explanation for the invariant association of the pCS.7 allele 2 with CF is that the mutation event that gave rise to CF in the Caucasian population occurred after racial divergence in man on a chromosome with haplotype B. These data argue against genetic heterogeneity at the CF locus, and we propose that the great majority of CF mutations found in the population arose from one single mutational event.

A close association has been shown previously between RFLP haplotypes at the β -globin locus and specific β -thalassaemia mutations³³, and a similar observation has been made for phenylketonuria³⁴. However, in some homogeneous populations, specific mutations have been observed almost exclusively associated with particular RFLPs, as for sickle-cell disease and the *HpaI* polymorphism³⁵, and for the β^0 -thalassaemia gene in Sardinia and the *BamHI* polymorphism³⁶.

Although the relationship between physical distance and genetic distance is not linear, as meiotic recombination between mammalian chromosomes is initiated at specific sequences ('hot spots') which are distributed non-uniformly³⁷, the relationship between disequilibrium and physical distance has been extensively studied³⁸. As a rule, linkage disequilibrium becomes stronger as physical distance and recombination rate between two loci decrease. If we assume a uniform recombination rate for 7q22–31 (the mild disequilibrium observed between pairs of loci in the *J3.11* and *met* interval on CF chromosomes¹³ argues against the presence of a recombinational hot spot) and selective neutrality of *XV-2c* and *CS.7*, we may calculate the effective recombination frequency between this pair of markers which we have shown to be 15 kb apart.

The disequilibrium coefficient (Δ) is related to the physical (d) and genetic (k) distance between the loci, and the effective population size (N_e): $\Delta^2 = 1/(1 + 4N_e k d)$. If N_e is assumed to 10,000, and for *XV-2c* and *CS.7* the Δ is –0.54 (on normal chromosomes) and physical distance (d) is 15 kb; then the recombination rate is 0.004 per Mb. For the pair of loci CF and *CS.7*, Δ is –0.62, N_e is 10,000, the recombination frequency is 0.000039, and the physical distance is calculated to be 9.8 kb. Similar results are found with a range of N_e values, or using

XV-2c instead of *CS.7* against CF. The population genetic data presented here suggest that *CS.7* and *XV-2c* both map to within 10 kb of CF.

The genetic data presented here have important implications for prenatal diagnosis and for carrier exclusion (as against carrier detection) in the general population. Almost 95% of the carriers for CF studied so far (70 persons in all) have at least one chromosome with haplotype B (*XV-2c* allele 1, *CS.7* allele 2). This haplotype (the CF haplotype in 94% of chromosomes) occurs in only 34% of the normal population (Table 2). Therefore, individuals with haplotypes A and C, whether homozygous or heterozygous, have a considerably reduced risk of being carriers for CF as compared to the 1 in 20 average risk in the British population; this includes >55% of the normal chromosomes. On the other hand, a homozygote for the haplotype B has an approximately one-seventh risk of being a carrier for CF. The association of CF to pCS.7 allele 2 may, after confirmation in a larger population sample, allow accurate carrier exclusion on its own.

Coding sequences

The fact that HTF islands usually identify a structural gene, and the strong linkage disequilibrium between CF and *pXV-2c* and pCS.7 (with no recombinants for pCS.7), suggest that the CF gene could be in the cNX.4-cNX.2 contig. A pool of single-copy fragments flanking the putative HTF island and *pXV-2c* was hybridized to Southern blots of *HindIII* or *EcoRI* digests of DNA from chicken, mouse, rat, rabbit, ox, sheep, pig, horse, marmoset and human (zoo-blot), and conserved sequences were found. Individual single-copy fragments were then used in a second hybridization with multiple zoo-blot panels. Most subclones hybridized to marmoset DNA, but only three subclones showed hybridization to all mammalian DNA samples tested and to chicken DNA (Fig. 5). These clones, pCS.7, *pXV-L3* and *pXV-2a*, are within a continuous region of 15 kb starting in pCS.7 (Fig. 2a), the clone that identifies the non-methylated sequence and that shows the highest disequilibrium with CF.

One cosmid from within the contig, cH.147, which included pCS.7, *pXV-L3* and *pXV-2a*, was used to screen a human lung complementary DNA library and Northern blots. The sequence

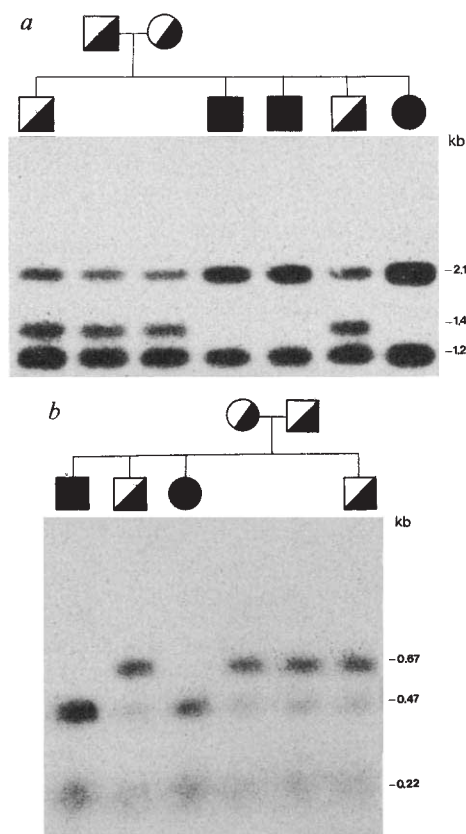


Fig. 4 Segregation of the allelic systems detected for pXV-2c and pCS.7. *a*, Probing with pXV-2c reveals a diallelic polymorphism (A1, 2.1 kb; A2, 1.4 kb) and a constant band of 1.2 kb with *TaqI*. *b*, Probing with pCS.7 detects a two allele polymorphism (A1, 0.67 kb, A2, 0.47 kb) and a constant band of 0.22 kb with *HhaI* (a methylation sensitive restriction enzyme). Square, male; circle, female; solid symbols, CF individuals. DNA was isolated from lymphocytes of family members and 5 μ g digested with *TaqI* or *HhaI*, electrophoresed and transferred to nylon filters. The 1-kb ladder was used as a size marker (now shown). Radiolabelling, hybridization, washing and autoradiography were as in Fig. 1.

is expressed in normal lung, placenta, intestine and kidney, and in CF placenta; a low level of expression was detected in brain RNA. This pattern of expression (data not shown) corresponds to the CF pathology. Positive lung cDNA clones have been obtained, mapped to the cNX.4-cNX.2 contig, and these are being investigated as candidate genes for CF.

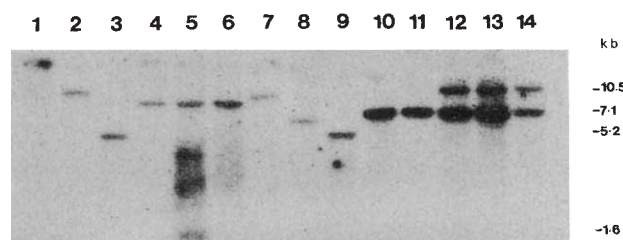


Fig. 5 Nucleotide sequence conservation across species of the cNX.4-cNX.2 contig. Different DNA samples (5 μ g) digested with *EcoRI* were hybridized with pCS.7 (a 2.0-kb *HindIII/NotI* fragment). Lane 1, chicken; 2, mouse; 3, rat; 4, rabbit; 5, ox; 6, sheep; 7, pig; 8, horse; 9, marmoset; 10, human (CF); 11, human; 12, clone 21; 13, clone 36; 14, CIId. A 1-kb ladder was used as size marker (not shown). The Southern blot was hybridized as described above; washing conditions were at 65 $^{\circ}$ C for 1 h in 0.2 $\times$ SSC and 0.1% SDS. Positive hybridization was also obtained using the subclones pXV-2a and pXV-L3 (data not shown). Other subclones from this contig (including pXV-2c) gave no cross-species homology. Mammalian DNAs were from R. Dodd; chicken DNA from S. John.

We have demonstrated, in the case of cystic fibrosis, the usefulness of selecting non-methylated regions as markers for expressed sequences and their power when used in conjunction with other physical techniques. This combination of genetic and molecular techniques for the identification of candidate genes should be of value for most genetic diseases in which the altered protein is unknown. For diseases for which linkage to a chromosome or chromosomal region has been found, this method represents an effective way of moving from detecting linkage to isolating the gene.

The Cystic Fibrosis Genetics Research Group is supported by the Cystic Fibrosis Research Trust (UK) and the MRC. We thank our many clinical and scientific colleagues for their help with probes, vectors, cell lines and techniques, and to the families for giving many blood samples and much encouragement. We thank Colin Cooper, Charles Coutelle, Charles Buys, Albert de la Chapelle, Kay Davies, Peter Goodfellow, Dickey Halley, Hans Lehrach, Inge Liebaers, Peter Little, Linda Meredith, Marianne Schwartz and Ray White. X.E., on leave from Hospital de Sant Pau of Barcelona, Catalunya, is supported by the British Council and the Spanish Ministry of Education and Science. M.F. is a Foulkes Foundation Fellow.

Note added in proof: We have analysed an additional CF family (GH, from Charles Coutelle, Berlin-Buch) which shows a maternal recombination between *met* and CF. The XV-2c *TaqI* polymorphism is informative in this family, and does not recombine with CF.

Received 18 March; accepted 1 April 1987.

- Royer-Pokora, B. *et al.* *Nature* **322**, 32-38 (1986).
- Monaco, A. P. *et al.* *Nature* **323**, 646-650 (1986).
- Friend, S. H. *et al.* *Nature* **323**, 643-646 (1986).
- Goodchild, M. C. & Dodge, J. A. *Cystic Fibrosis* (Bailliere Tindall, London, 1985).
- Eiberg, H., Mohr, J., Schmiedel, K., Nielsen, L. S. & Williamson, R. *Clin. Genet.* **28**, 265-271 (1985).
- White, R. *et al.* *Nature* **318**, 382-384 (1985).
- Wainwright, B. J. *et al.* *Nature* **318**, 384-385 (1985).
- Tsui, L.-C. *et al.* *Science* **230**, 1054-1057 (1985).
- Scambler, P. J. *et al.* *Lancet* **ii**, 1241-1242 (1985).
- Scambler, P. J. *et al.* *Nucleic Acids Res.* **14**, 1951-1956 (1986).
- Wainwright, B. J. *et al.* *Am. J. hum. Genet.* (submitted).
- Estivill, X., Schmidtke, J., Williamson, R. & Wainwright, B. J. *Hum. Genet.* **74**, 320-322 (1986).
- Beaudet, A. *et al.* *Am. J. hum. Genet.* **39**, 681-693 (1986).
- Lathrop, G. M. *et al.* *Science* (in the press).
- Scambler, P. J., Law, H.-Y., Williamson, R. & Cooper, C. S. *Nucleic Acids Res.* **14**, 7159-7174 (1986).
- Scambler, P. J. *et al.* *Hum. Genet.* (in the press).
- Scambler, P. J. *et al.* *Am. J. hum. Genet.* (submitted).
- Wainwright, B. J. *et al.* *Cytogenet. Cell Genet.* (in the press).
- Scambler, P. J. *et al.* *Nucleic Acids Res.* (in the press).
- Estivill, X. & Williamson, R. *Nucleic Acids Res.* **15**, 1415-1425 (1987).
- Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821-7825 (1986).
- Smith, C. L., Warburton, P. W., Gaal, A. & Cantor, C. R. *Genet. Engn.* **8**, 45-70 (1986).
- Brown, W. R. A. & Bird, A. P. *Nature* **322**, 447-481 (1986).
- Bird, A. P. *Nature* **321**, 209-213 (1986).
- McClelland, M. & Nelson, M. *Nucleic Acids Res.* **13**, r201-207 (1985).
- Kolsto, A.-B. *et al.* *Nucleic Acids Res.* **14**, 9667-9678 (1986).
- Croce, C. M. & Kaprowoski, H. J. *exp. Med.* **139**, 1350-1353 (1974).
- Knowles, B. B. *et al.* *J. exp. Med.* **145**, 314-326 (1977).
- Michiels, F., Burmeister, M. & Lehrach, H. *Science* (submitted).
- Park, M. *et al.* *Cell* **45**, 895-904 (1986).
- Farrall, M. *et al.* *Am. J. hum. Genet.* **39**, 713-719 (1986).
- Hill, W. G. & Robertson, A. *Theor. appl. Genet.* **38**, 226-231 (1968).
- Kazazian, H. H. *et al.* *Nature* **310**, 152-154 (1984).
- DiLella, A. G., Marvit, J., Lidsky, A. S., Guttler, F. & Woo, S. L. C. *Nature* **322**, 799-803 (1987).
- Kan, Y. W. & Dozy, A. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5631-5635 (1978).
- Kan, Y. W., Lee, K. Y., Furbeta, M., Angius, A. & Cao, A. *New Engl. J. Med.* **302**, 185-188 (1980).
- Steinmetz, M., Uematsu, Y. & Lindahl, K. F. *Trends Genet.* **3**, 7-10 (1987).
- Chakravarti, A. *et al.* *Am. J. hum. Genet.* **36**, 1239-1258 (1984).
- Cross, S. H. & Little, P. F. R. *Gene* **49**, 9-22 (1986).
- Grosfeld, F. G. *et al.* *Gene* **13**, 227-237 (1982).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Poustka, A., Pohl, T. M., Barlow, D. P., Frischauf, A.-M. & Lehrach, H. *Nature* **325**, 353-355 (1987).
- Carle, G. F. & Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3756-3760 (1985).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).

Structure of the repressor–operator complex of bacteriophage 434

J. E. Anderson*, M. Ptashne & S. C. Harrison

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

The crystal structure of a specific complex between the DNA-binding domain of phage 434 repressor and a synthetic 434 operator DNA shows interactions that determine sequence-dependent affinity. The repressor recognizes its operators by its complementarity to a particular DNA conformation as well as by direct interaction with base pairs in the major groove.

WE describe here the structure of a specific complex between the DNA-binding domain of a bacteriophage repressor protein and a synthetic operator DNA. The description is based on an X-ray crystallographic analysis of crystals that diffract to 3.2 Å resolution in some directions and to about 4.5 Å in others. The repressor is encoded by coliphage 434, a close relative of phage λ . It binds to a set of six similar but non-identical 14-base-pair (bp) operators in the phage genome. Its differential affinity for these sites, coupled with the distinct differential affinity of the homologous cro protein, creates a regulatory switch determining the choice between lysogeny and lytic growth¹. The structure of the repressor–operator complex, taken together with experiments reported in an accompanying paper and elsewhere^{2,3}, reveals the basis of specific binding and its modulation by the exact DNA sequence of the operator.

In an account of this structure at 7 Å resolution⁴, we described the general features of the complex. The repressor DNA-binding domain (R1-69), contains the first 69 residues of the complete polypeptide. Its conformation is predominantly α -helical, similar to the first four helices of the λ repressor⁵. The second and third alpha helices ($\alpha 2$ and $\alpha 3$) form a helix–turn–helix motif, found in crystal structures of certain other DNA binding proteins^{6–8} and believed on the basis of amino-acid sequence similarity to occur in many more⁹. The conformation of the 14-bp synthetic operator (14-mer) is B-DNA-like, and individual 14-mers are stacked end-to-end to form pseudocontinuous double helices running through the crystals. Two R1-69 subunits are bound to each 14-mer, and $\alpha 3$ of each monomer rests in the major groove of a half-site.

The higher resolution structure presented here shows the interaction in molecular detail. Hydrogen bonds between amide nitrogens of the peptide backbone and phosphate groups on DNA strongly constrain the fit of DNA to protein, and three glutamine side chains on $\alpha 3$ are in position to form a pattern of van der Waals and hydrogen-bond contacts with the five outer base pairs. Thus, the repressor recognizes in DNA both a particular conformation and an array of base-pair contacts.

Structure determination

Crystals of R1-69 and the 14mer were prepared as described previously¹⁰. Data-collection procedures and statistics are presented in Table 1. Initial single isomorphous replacement (SIR) phases were obtained using crystals with DNA containing 5-bromodeoxyuracil instead of thymine at position 7. Phases between 15 Å and 7 Å were refined using real space noncrystallographic symmetry averaging^{11–13} about a local 3₁ axis parallel to [111] and intersecting [110] 33.5 Å from the origin⁴. The small magnitude of the heavy-atom differences and weak intensities beyond 7 Å made conventional refinement of heavy-atom positions impractical. We therefore used a two-step strategy to extend phases, first from 7 Å to 5 Å and then from

5 Å to 3.2 Å. In the first step, heavy-atom positions were refined at 7 Å resolution by averaging a 15–7 Å heavy-atom difference map about a series of local 3₁ axes, parallel to [111] and intersecting [110] in the neighbourhood of 33.5 Å from the origin. The position of the 3₁ axis that gave the strongest average heavy-atom peaks was selected, and new heavy-atom positions were determined by inspection of the corresponding map. SIR phases between 7 and 5 Å resolution were computed using these positions, combined with the previously refined 15 to 7 Å phases, and used to generate a 5 Å map. Eight cycles of noncrystallographic symmetry averaging¹³ about the new local 3₁ (34.0 Å from the origin along [110]) gave an averaging R -factor, $R_{av} = 0.28$. In the second step, refined phases to 5 Å resolution were used in an analogous way to adjust the origin of the threefold screw axis and to improve the heavy-atom coordinates. The noncrystallographic symmetry phase refinement was unstable if all phases from 15 to 3.2 Å were allowed to change, but fixing phases from 15 to 5 Å at their best values from the 5 Å refinement allowed the computation to converge with $R_{av} = 0.30$ for all data from 15 to 3.2 Å (Table 2, upper part). This way of refining heavy-atom coordinates is similar to the procedure introduced by Hogle for work on poliovirus¹⁴.

The map thus computed showed clear double-helical DNA density, each deoxyribose-phosphate backbone having a staircase like shape, with sugars corresponding to the risers. Base pairs were poorly defined by density, but an unambiguous model could be built using backbone density and hydrogen-bonding restraints as a guide. Four α -helices of the protein appeared as twisted rods. A number of side chains were well-defined; others appeared as truncated bulges. For model-building, we used the program FRODO¹⁵ with separate files for DNA and protein coordinates. The 14mer was broken into single stranded mono-, di- and trinucleotide segments, the phosphates and sugars were fitted to the DNA backbone density, and the torsion angles were adjusted to restore approximately correct base-pair geometry. The Hendrickson–Konnert restrained least squares refinement programs¹⁶, modified by G. Quigley (MIT) to accept nucleic acid coordinates, were used with base-pair hydrogen bonding restraints to restore idealized geometry. Further manual adjustment followed, with periodic regularization. Residues 1–58 of the protein were built into the map and adjusted with the regularization routines in FRODO. Finally, protein and DNA coordinates were regularized with the constrained-restrained least squares refinement program CORELS¹⁷, with base-pair hydrogen bonding restraints imposed on the DNA.

The initial model, constructed as just described, was used to obtain starting phases for a second round of noncrystallographic symmetry phase refinement. The order in the crystals is anisotropic, with intensities measurable to spacings of 3.2 Å in the directions of noncrystallographic symmetry axes formed by stacked protein–DNA complexes, but only to about 4.5 Å in other directions¹⁰. The map can therefore be regarded as density at 3.2 Å resolution, selectively ‘smeared’ in directions perpendicular to the non-crystallographic symmetry axis. When build-

* Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, Long Island, New York 11724, USA.

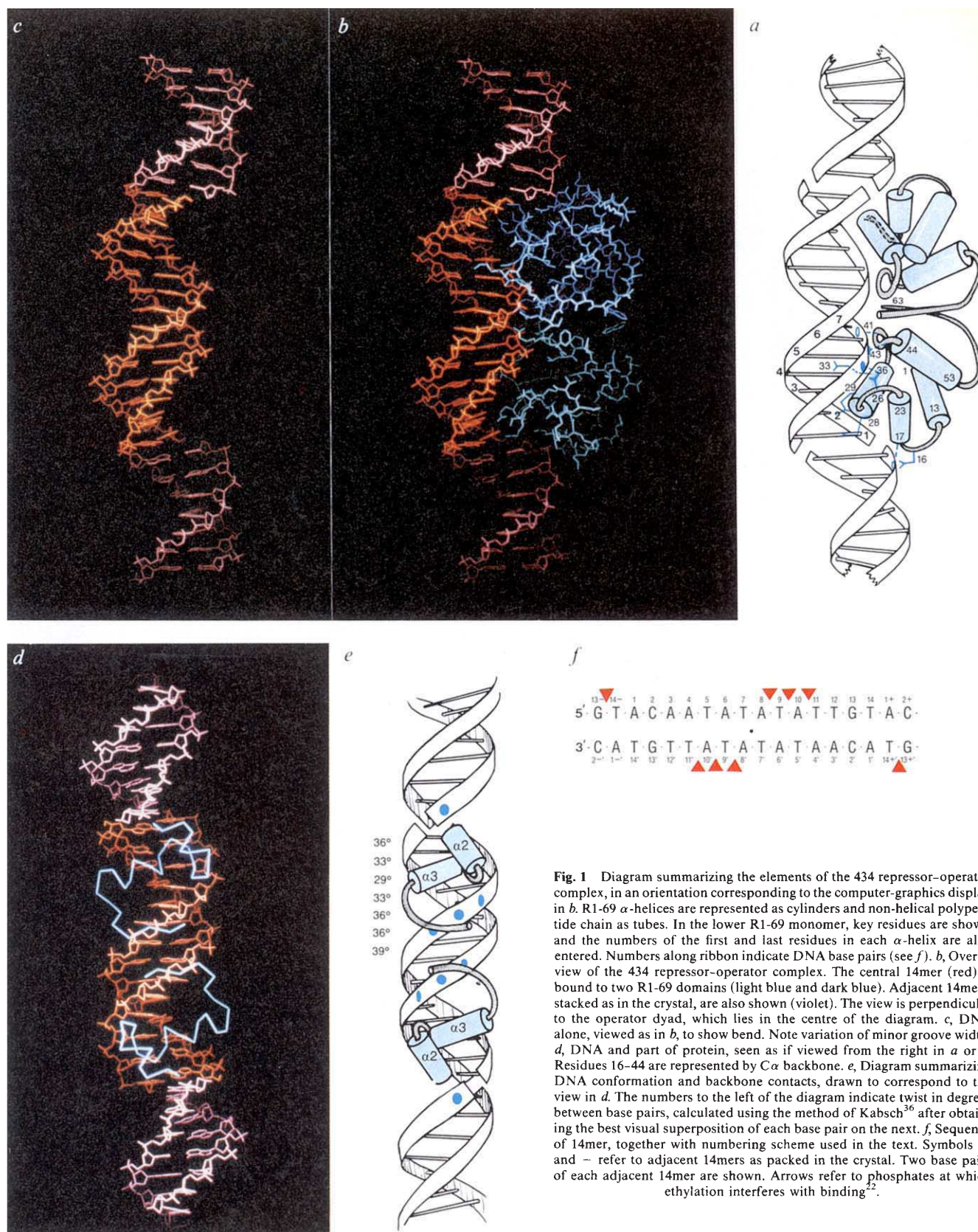


Fig. 1 Diagram summarizing the elements of the 434 repressor-operator complex, in an orientation corresponding to the computer-graphics display in *b*. R1-69 α -helices are represented as cylinders and non-helical polypeptide chain as tubes. In the lower R1-69 monomer, key residues are shown and the numbers of the first and last residues in each α -helix are also entered. Numbers along ribbon indicate DNA base pairs (see *f*). *b*, Overall view of the 434 repressor-operator complex. The central 14mer (red) is bound to two R1-69 domains (light blue and dark blue). Adjacent 14mers, stacked as in the crystal, are also shown (violet). The view is perpendicular to the operator dyad, which lies in the centre of the diagram. *c*, DNA alone, viewed as in *b*, to show bend. Note variation of minor groove width. *d*, DNA and part of protein, seen as if viewed from the right in *a* or *b*. Residues 16-44 are represented by $C\alpha$ backbone. *e*, Diagram summarizing DNA conformation and backbone contacts, drawn to correspond to the view in *d*. The numbers to the left of the diagram indicate twist in degrees between base pairs, calculated using the method of Kabsch³⁰ after obtaining the best visual superposition of each base pair on the next. *f*, Sequence of 14mer, together with numbering scheme used in the text. Symbols + and - refer to adjacent 14mers as packed in the crystal. Two base pairs of each adjacent 14mer are shown. Arrows refer to phosphates at which ethylation interferes with binding²².

ing a model, however, stereochemical constraints couple positions along the local symmetry axis with those normal to it. Therefore, the effective resolution is intermediate between the two limits. To 5 Å, the intensity fall-off is relatively isotropic and computed structure factors at this resolution were used in a translation search program (A. Aggarwal, personal communication) to carry out a final adjustment of the position of the 3₁

axis. The anisotropy and non-crystallographic symmetry were incorporated into FFT (fast Fourier transform) structure factor calculations¹⁸ by using the Bricogne density reconstitution procedure¹³ to generate a full crystallographic asymmetric unit from a half-complex 'map' that is computed in a coordinate frame with the noncrystallographic symmetry axis along Z. This choice of frame simplified introduction of a uniaxial, anisotropic 'tem-

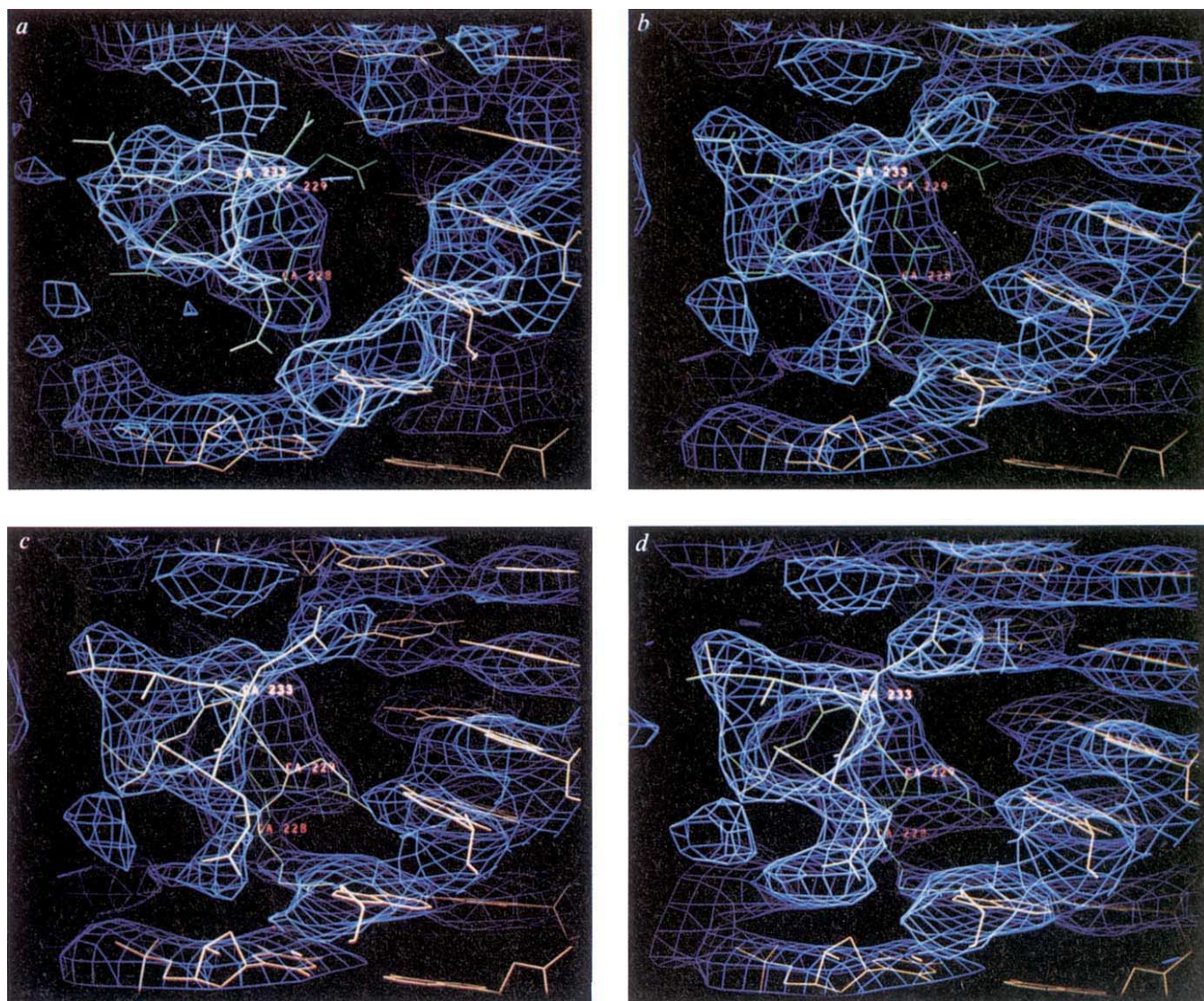


Fig. 2 Part of the electron density maps at different stages of the structure determination, showing the N-terminal half of $\alpha 3$ and portions of base pairs 1-4. Models are superimposed on density. The helix $\alpha 3$ is viewed from the C-terminal end, with residues 28-34 (QQSIEQL) shown and α -carbons of residues 28, 29 and 33 labelled (228, 229 and 233, respectively). *a*, Map after refinement of SIR phases, with initial model superimposed. *b*, Map after refinement of phases determined from initial model, with initial model superimposed. Note clear indications for repositioning of side chains, especially Gln 28 and Gln 29 (background) and Gln 32 (foreground). *c*, Same map as in *b*, but with rebuilt model. *d* Final ($2F_o - F_c$) map, with rebuilt model.

perature factor'. Various ratios of $B_x = B_y$ to B_z were examined, by comparing calculated and observed amplitudes. The optimum occurred at 1.75, as judged both by R factor minimum and by relative lack of noise in corresponding $2F_o - F_c$ and $F_o - F_c$ maps.

Phases from the initial model were used to initiate eight further cycles of noncrystallographic symmetry refinement of phases between 15 and 3.2 Å. The computation converged stably to $R_{av} = 0.27$, with an overall phase change of 43° (Table 2, lower part). Significant adjustments were indicated by this map, and we extended the model from residue 58 to residue 63 (Fig. 1). Two further rounds of smaller adjustments were based on $2F_o - F_c$ and $F_o - F_c$ maps. Maps and coordinates at various stages are shown in Fig. 2. It is evident that the second round of non-crystallographic-symmetry refinement substantially improved phase determination. Base-pair and side-chain densities became far clearer and peptide-backbone density more continuous. Poor placement of some side chains of the initial model did not prevent appearance of density in correct positions in the final map, showing that errors in that model did not unduly bias the outcome of the phase refinement (Fig. 2*b* and *c*). The use of an intermediate model to initiate a second round

of phase refinement was used in determining the turnip crinkle virus structure, where similar corrections to an actual structure were generated¹⁹. The power of threefold non-crystallographic redundancy in phase refinement has been amply demonstrated in previous work—for example, in the determination of the influenza virus haemagglutinin structure, with data of comparable accuracy to ours²⁰. Reciprocal-space refinement of the model is in progress. The R -factor of the model presented here is 0.44. (H. Holley, J.E.A. and S.C.H. unpublished data).

As in any structure determination at this resolution, interactions such as hydrogen bonds must be inferred from the position and orientation of the participating groups. We believe our present structure to be sufficiently well determined that the protein-DNA interactions can be described correctly and that likely hydrogen bonds can be assigned. More precise conformational details, accurate hydrogen-bond geometry, contact distances and features such as the positions of water molecules—all of which may be significant for complete understanding of specificity—will only be visible at the higher resolution afforded by a different R1-69-operator crystal currently being studied (A. Aggarwal and S.C.H. unpublished data).

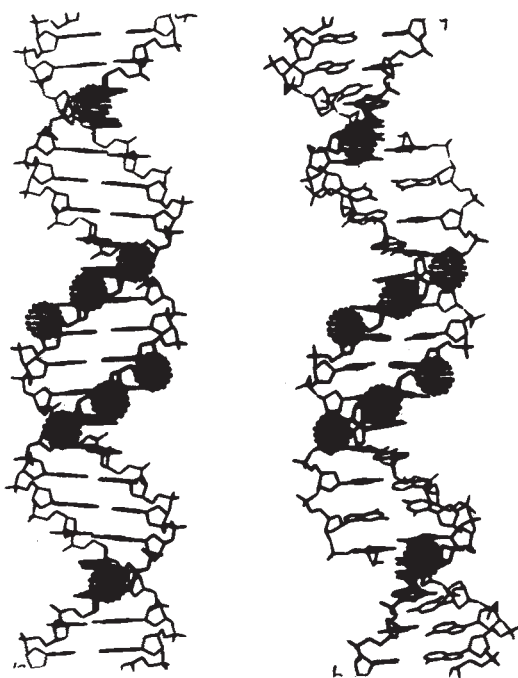


Fig. 3 Conformation of DNA in the crystalline complex (right) compared with idealized, 10.5-bp-per-turn B-DNA (left). Contacted phosphates are highlighted. It is clear from this figure, and from Fig. 1, that model building with idealized B-DNA could not correctly predict the protein-DNA contacts.

DNA conformation

The DNA of the 14mer forms a B-type helix throughout its length. The 14mers are stacked accurately on adjacent DNA segments to form a pseudocontinuous helix. The local twist varies from about 39° to 29° per base pair—that is, from 9.2 to 12.3 base pairs per turn (Fig. 1e). A similar range of variation in twist has been found in crystalline B-DNA²¹. With respect to the average of 10.5 base pairs per turn, the 14mer is overwound at its centre and underwound at its ends. The helix axis bends somewhat, causing the DNA to curl slightly around helix 3 of each monomer. The bend is sharpest between base pairs 4 and 5 (and 10 and 11), with relatively more gentle bending between base pairs 1 and 4, between base pairs 5 and 10, and between base pairs 11 and 14. Overwinding and bending in the centre of the DNA fragment narrows the minor groove. Underwinding and bending widens it at the ends. This variation is evident in Figs 1 and 3.

Protein conformation

R1-69 is a cluster of four α -helices, with a C-terminal extension. Two of the helices, α 2 and α 3, form a helix-turn-helix motif. The residues found in each of the four helices are shown in Table 3. For comparison, the corresponding residues in the N-terminal domain of λ repressor⁵, determined by comparison of the two models, are also listed. As observed in our earlier paper⁴, if the 434 and λ α 2- α 3 structures are superposed, the axes of α 1 and α 4 in the two repressors coincide to within about 2–3 Å. The α -carbon positions suggest that 434 repressor has 'deletion' with respect to λ repressor of two residues between α 1 and α 2 and of one residue at the beginning of α 4. The two models diverge at residue 58 of 434 (corresponding to 75 of λ), precisely the point at which the two amino-acid sequences cease to be similar. Various hydrophobic side-chains form the interior of the four-helix cluster, which is also stabilized by several polar linkages—notably, Gln 17–Glu 32 and Arg 5–Glu 35. The last six residues of R1-69 are poorly defined in our map. They appear to extend toward DNA backbone near P11.

Interactions occur between the two R1-69 subunits bound to one 14mer. The two monomers are in contact across the dyad

Table 1 Statistics for data between 45 and 3.2 Å resolution

Native				
$d_{\min}$ (Å)	$\langle F \rangle$	$R_{\text{sym}} (N_{\text{mult}})$	N (% $F > 2\sigma$)	N_{poss} (% $F > 2\sigma$)
9.90	7,568	0.080 (148)	368 (98.4)	622 (58.2)
7.08	6,559	0.096 (585)	929 (98.5)	622 (58.2)
5.81	3,476	0.107 (721)	1,099 (97.0)	1,260 (84.6)
5.04	2,916	0.138 (815)	1,283 (95.9)	1,466 (83.9)
4.51	2,955	0.157 (920)	1,399 (94.1)	1,651 (79.8)
4.12	3,022	0.170 (951)	1,535 (94.2)	1,817 (79.6)
3.82	2,695	0.212 (982)	1,675 (92.0)	1,961 (78.6)
3.58	2,491	0.242 (941)	1,705 (88.7)	2,090 (72.3)
3.37	2,161	0.284 (921)	1,756 (82.8)	2,237 (65.0)
3.20	2,535	0.212 (834)	1,708 (79.7)	2,328 (58.5)
Overall	3,129	0.161 (7,818)	13,457 (90.7)	16,444 (74.2)
Derivative				
$d_{\min}$ (Å)	$\langle F \rangle$	$R_{\text{sym}} (N_{\text{mult}})$	N (% $F > 2\sigma$)	N_{poss} (% $F > 2\sigma$)
9.90	7,194	0.041 (59)	229 (97.4)	622 (35.9)
7.08	6,225	0.047 (240)	664 (96.8)	1,013 (63.5)
5.81	3,341	0.093 (387)	866 (88.6)	1,260 (60.9)
5.04	2,879	0.135 (442)	970 (84.7)	1,466 (56.1)
4.51	2,902	0.179 (471)	1,079 (85.1)	1,651 (55.6)
4.12	2,905	0.212 (424)	1,132 (83.7)	1,817 (52.2)
3.82	2,790	0.282 (366)	1,244 (79.7)	1,961 (50.6)
3.58	2,630	0.313 (274)	1,200 (75.5)	2,090 (43.3)
3.37	2,430	0.274 (244)	1,141 (70.9)	2,237 (36.2)
3.20	3,405	0.194 (261)	1,089 (75.0)	2,328 (35.1)
Overall	3,224	0.169 (3,168)	9,614 (81.6)	16,444 (47.7)

Mean isomorphous difference = $\sum |F_{\text{nat}} - F_{\text{der}}| / \sum |F_{\text{nat}}| = 0.22$, calculated after exclusion of data (between 45 and 3.2 Å) for which $|F_{\text{nat}} - F_{\text{der}}| > 5\sigma(\Delta F)$.

$$R_{\text{sym}} = \frac{\sum_i \sum_j |I_{ij} - \bar{I}_i|}{\sum_i \sum_j I_{ij}}$$

N_{mult} = number of reflections measured more than once; only these contribute to R_{sym} .

N = number of reflections measured (after rejections) including those measured only once.

N_{poss} = maximum number of reflections possible in corresponding resolution range.

The percentage of observed and possible data with $F > 2\sigma(F)$ is indicated. Data from native crystals were collected to 3.2 Å using CEA-25 X-ray film and oscillation photography with Elliot GX-6 and GX-13 X-ray generators. The films were scanned with an Optronics P1000 film scanner on a 50 μ m raster. Integrated intensities were obtained with the film-scanning program SCANFILM, a derivative of SCAN12 (ref. 34). Data to 7 Å were also collected on a prototype Xentronics area detector with the Harvard software package³⁵. These measurements provided intensities for native reflections that were too bright to be measured on film. Derivative data to 3.2 Å were collected on the area detector from isomorphous crystals prepared with DNA containing 5-bromodeoxyuracil instead of thymine at position 7 of the 14mer⁴. The native data from detector and film were processed and merged, then scaled to the processed derivative data. All programs for data processing and scaling were from P. Evans (MRC Laboratory of Molecular Biology).

near residues 57–58. The aliphatic chain of Arg 41 and a residue near the C-terminus in one subunit lie against the aromatic ring of Phe 44 in the other subunit. These contacts may be important in determining the relative orientations of the DNA-binding surfaces of the two monomers, thereby defining the required spatial orientation of one operator half-site with respect to the other.

Protein-DNA contacts

The polypeptide chain is so folded that in the complex the N-termini of α -helices 2, 3 and 4 all point toward the DNA backbone, α 3 lies in the major groove, and the loop joining α 3 and α 4 runs along the DNA backbone (Fig. 1d). Contacts to DNA backbone from one R1-69 subunit occur on both sides of the major groove that is occupied by α 3 and possibly across the minor groove that contains the operator dyad. Major-groove interactions with base pairs 1–5 and 10–14 are made by three glutamine side chains projecting from α 3. These relationships are summarized in Figs 1d and e, and described in detail below. **Backbone contacts.** Four phosphate groups per half-site, corresponding precisely to the phosphates implicated by ethylation interference experiments²², appear to interact with protein. Considering half of the complex, the contact phosphate groups lie

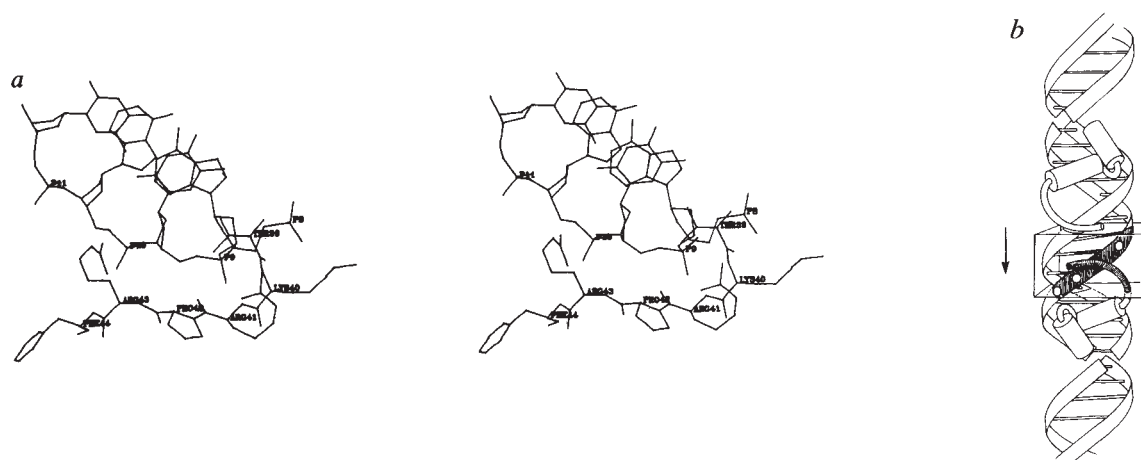


Fig. 4 Stereo view of contacts between protein and DNA phosphates P9' and P10'. b, Schematic diagram as a key to a. The view is of the region indicated in the (shaded structures), looking approximately along the DNA axis (arrow). Compare key with Fig. 1e. Residues 39–44 of the protein are shown, together with nucleotides 8'–11'. (Primes are omitted in the labels.)

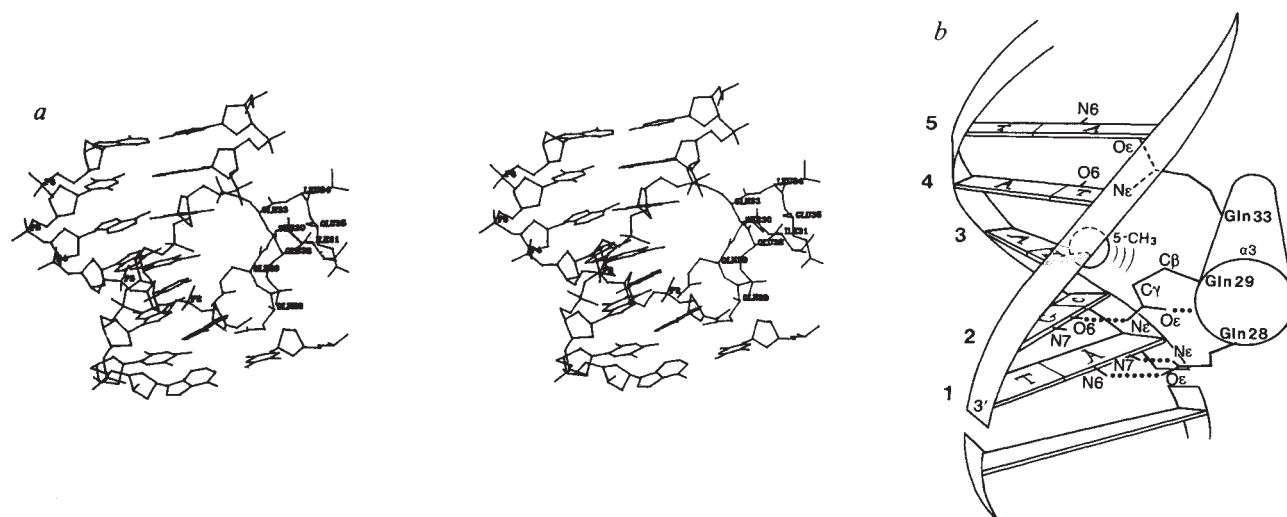


Fig. 5 a, Stereo view of helix $\alpha 3$ and base pairs 1–5, viewed from the N-terminus of $\alpha 3$. One base pair of the adjacent 14-mer is also present. Included are residues 28–35 of R1-69. b, Schematic diagram as a key to a. Note the similarity to view of $\alpha 3$ and surrounding structure in the lower monomer of Fig. 1c. Gln 28, Gln 29 and Gln 33 are shown, with hydrogen bonds suggested by the current model drawn in dotted lines. Functional groups on bases and on Gln 33 that might participate in additional hydrogen bonds are also shown. The proposed non-polar contact between the 5-CH₃ group of T12' (base pair 3) and C β and C γ of Gln is indicated.

5' to bases 14-, 9', 10' and 11' (see Fig. 1f for numbering scheme). The P14- contact occurs between different complexes in the crystal. Because the stacking of 14mers mimics continuous DNA, we believe that the interaction we observe at P14- closely reflects the one actually made in continuous DNA.

The three tightest contacts (as judged both from the structure and by the strength of ethylation interference) all appear to involve peptide backbone NH groups. At P14-, the NH of Gln 17 approaches the phosphate oxygens closely enough to form a hydrogen bond. Residue 17 is at the N-terminus of $\alpha 2$, and its peptide NH group is expected to make a particularly strong hydrogen bond to a negatively charged ion, as a result of oriented peptide dipoles in the α -helix²³. Analogous interactions occur in the *Salmonella typhimurium* sulphate binding protein, where peptide NH groups at the N-termini of α helices are ligands of bound SO₄²⁻ ions²⁴. The side chain of Asn 16 also appears to be involved in the contact to P14-. At P9', there is close approach of the main chain NH groups of Lys 40 and Arg 41; the side-chain conformation of Arg 41 permits a salt link with the same phosphate (Fig. 4). At P10', there is close approach of the peptide NH groups of both Arg 43 and Phe 44 (Fig. 4). Phe 44 is the amino-terminal residue of $\alpha 4$. The use of peptide NH groups to contact phosphates may be a common feature in proteins that bind DNA and nucleotides. Dreusicke and Schulz²⁵ have

described a 'giant anion hole' at the site in adenylate kinase where a nucleotide phosphoryl group is believed to be located. The polypeptide chain loops in such a way that five peptide NH groups coordinate a bound sulphate in their crystal structure.

The side chain of Arg 43 projects into the minor groove near base pair 7. The groove is somewhat narrower than average at this position, and the guanidinium group could probably have significant coulombic interaction with phosphates 9 and 10' on either side of the groove (Fig. 4). The possibility of direct interaction with base pair 7 is discussed in the next section and in the accompanying paper.

The ethylation of P11 interferes more weakly with binding than ethylation of P14-, P9' or P10' (ref. 22). Thr 26 appears to contact P11', and the C-terminal residues of the other monomer may also lie in this vicinity. Because the C-terminal residues are not well defined in our map, we cannot at present determine whether any actually engage in backbone contacts. If they do, the interaction may help establish the favoured relative orientation of the two operator half-sites, because an R1-69 domain bound in one half-site would make contact by its C terminus with DNA backbone in the other half-site.

Base-pair contacts. Three glutamines (residues 28, 29 and 33) project from $\alpha 3$ into the major groove (Fig. 5). Gln 28

approaches A1 (base pair 1) in such a way that hydrogen bonds can readily form between N7 of the base and N ϵ of the glutamine and between N6 and O ϵ . 'Bidentate' H-bonding of this type was suggested as a mode of adenine-specific interaction by Seeman *et al.*²⁶ Gln 29 lies along α 3. Its C β and C γ are in van der Waals contact with the 5-methyl group of T12' (base pair 3), and its N ϵ can hydrogen bond to O6 (and perhaps also N7) of G13' (base pair 2). The O ϵ of Gln 29 may hydrogen bond to its main chain NH, an interaction occasionally seen with glutamine at amino termini of α -helices²⁷. This hydrogen bond would fix the orientation of the glutamine and point the N ϵ donor group directly at G13'. Gln 33 projects toward T11' and A10' (base pairs 4 and 5). There could be a hydrogen bond between N ϵ of Gln 33 and O4 of T11' (base pair 4), and perhaps between O ϵ and N6 of A10' (base pair 5); a refined model will be required for confident assignment.

The present model shows no direct interactions with base-pairs 6 and 7. Although residues 64–69 are not well defined in the map they are clearly excluded from the major groove on the near side of the operator by α 3 and from the minor groove by residues 40–44. Moreover, model building indicates that there is no reasonable possibility of base-pair contact from these residues in the major groove on the far side of the DNA.

Contributions to specificity

Three different features of the DNA structure appear to be significant for specificity: (1) the groups accessible in the major groove on base pairs 1–5, (2) the conformational details of the half-sites, and (3) the overwinding of the central base pair steps with a narrowing of the minor groove. These features are relevant to different classes of interactions made by the protein. (1) Side chains from α 3 contact base pairs in the major groove. (2) Peptide amino groups, as well as side chains from various residues, contact backbone phosphates in ways that constrain the conformation of the operator half-sites. (3) Dimer interactions between repressor monomers favour a particular spatial relationship between half-sites.

Base-pair contacts. The interactions of Gln 28 with base pair 1 and of Gln 29 with base-pairs 2 and 3 are consistent with the conservation of base pairs 1–3 in all 12 operator half-sites, because changes in any of these base pairs would destroy one or more of the hydrogen bonds or disturb van der Waals contacts with the thymine methyl group. (For a list of all naturally-occurring operator sequences, see accompanying paper, ref. 2.) Wharton³ used a random mutagenesis scheme to generate all possible changes in amino acids 28 and 29. No viable mutants were detected, and indeed no likely substitutions in the model can be imagined. Wharton and Ptashne²⁸ have also shown that a mutant operator with T instead of A at position 1 is recognized specifically by a mutant repressor with alanine at position 28. These coordinated changes can be built into the model, generating a van der Waals contact between the $-\text{CH}_3$ group of T1 and the $-\text{CH}_3$ side chain of Ala 28.

The side chain of Gln 33 is in position to contact base pairs 4 and 5. The model shows that this residue could hydrogen bond to either base 11' or 10', or to both (Fig. 5). Wharton³ has studied the effects of substituting Ala for Gln 33 on the binding of repressor to various operators. His results, taken with others (R. Wharton and M.P. unpublished data), are consistent with the idea that residue 33 indeed contacts positions 4 and 5.

Backbone contacts. Within a half-site, contacts to the DNA backbone position a subunit of the protein against the operator. The relationship is likely to be quite exact, because main-chain peptide amino groups participate in the contacts and the main chain of a folded protein is generally less flexible than the side chains. The backbone contacts, especially the one at P14, require a gently bent outer operator²². This curvature is essential for at least one of the major groove contacts: without it, the interaction between Gln 28 and A1 would not have the correct geometry. It is not yet clear whether the bend is present in

Table 2 Phase refinement by noncrystallographic symmetry averaging

Refinement of SIR phases*				Phases refined 5 $\rightarrow$ 3.2 Å		
Cycle	R_{av}	r	Mean phase change	R_{av}	r	Mean phase change
1	0.58	0.537	53.3	0.55	0.346	75.3
2	0.45	0.734	40.1	0.41	0.600	34.7
3	0.39	0.785	21.9	0.35	0.678	17.1
4	0.38	0.791	14.5	0.33	0.718	10.2
5	0.38	0.780	11.4	0.31	0.737	7.3
6	0.39	0.780	10.2	0.31	0.750	5.7
7	0.40	0.781	10.5	0.30	0.758	4.8
8	0.43	0.747	11.0	0.30	0.785	4.1
Overall			65.5			82.0

Refinement of phases determined from initial model†

Cycle	R_{av}	r	Mean phase change
1	0.37	0.792	23.2
2	0.32	0.837	9.2
3	0.30	0.858	6.3
4	0.29	0.866	4.7
5	0.28	0.873	3.8
6	0.28	0.878	3.3
7	0.27	0.881	2.7
Overall			43.4

* SIR phases were refined by symmetry averaging maps calculated with coefficients $(2F_o - F_c) \exp(i\alpha_c)$. The molecular envelopes required for symmetry averaging¹³ were generated from model coordinates by enclosing each atom in a sphere of 7 Å radius. The initial model for the R1-69:14mer complex (Anderson *et al.* 1985), used to produce the envelope for 5 Å symmetry averaging (see text), contained idealized straight B-DNA. Inspection of the 5 Å map revealed that a 14mer with a slight bend would fit the density better. We therefore adjusted the positions of the two half-sites as rigid bodies using the real space refinement option of FRODO, and carried out the eight cycles of 3.2 Å refinement using an envelope from this model. Mean phase changes $(\sum|\phi_o - \phi_c|/N)$, for cycles subsequent to the first, are for all data present in the native data set, whether or not present in the initial SIR data set.

† Phases calculated from coordinates fit to the averaged maps from *a* were refined by averaging $F_o \exp(i\alpha_c)$ maps. The envelope was constructed from the coordinates on which the initial phases were based. $R_{av} = \sum|F_o - F_c|/\sum|F_o|$

$$r = \frac{\sum|F_o||F_c| - (\sum|F_o|)(\sum|F_c|)}{[\sum|F_o|^2 - (\sum|F_o|)^2] \cdot [\sum|F_c|^2 - (\sum|F_c|)^2]}$$

unbound operator DNA, or whether protein binding induces it. **Central base pairs.** The configuration of the operator at its central base pairs is consistent with sequence preferences at positions 6–9. A more detailed discussion appears in the accompanying paper, but we summarize the arguments here. The twist is about 36° between base pairs 6 and 7 and base pairs 8 and 9 and 39° between base pairs 7 and 8 (Fig. 1e). The minor groove is consequently 2.5 Å narrower than in idealized B-DNA. It is believed that GC or CG base pairs are not well accommodated in a narrow minor groove²⁹, and repressor indeed discriminates against GC or CG at positions 6–9 (ref. 2). The sequence preferences at central base pairs persist when alanine is substituted for Arg 43, which projects into the minor groove. The position of the guanidinium group in the present model does not indicate hydrogen bonding to bases, and the binding affinities of the Ala 43 mutant show that any minor-groove contacts that are made do not contribute significantly to sequence discrimination at positions 6–9. We therefore believe that it is the DNA conformation, rather than contacts to the bases themselves, that influences binding.

How does the structure in the centre of the operator influence affinity? One explanation is that a particular conformation is required for correct relative orientation of the two half-sites. Though monomeric at most concentrations in solution, the two R1-69 domains in the complex with DNA have significant interactions, including complementary packing of parts of each subunit that are important for DNA backbone contacts (especially residues 41–44; see Fig. 4). Thus, the relative orientation of

monomer DNA-binding surfaces is influenced by interactions between individual R1-69 subunits. The tightly structured repressor domain binds to a half-site in a precise way. Any variation in twist or bend at the central base pairs changes the spatial relationship of the two half-sites and requires compensatory variation in the protein dimer contacts, with a corresponding loss in the overall energy of the complex. We expect dimer contacts to be more extensive and less deformable in the intact repressor, and indeed alterations in central base-pair sequence have greater effects on the binding of the intact molecule than on R1-69 (ref. 2). Analogous mechanisms have been suggested on the basis of indirect evidence as a contribution to the specificity of SV40 T antigen³⁰.

The conformation of operator bound to a dimer of R1-69 is likely to differ from the equilibrium structure of the DNA in solution, but the base sequence will in any case affect flexibility and hence its propensity to reconfigure on binding. An even greater perturbation has been seen in the complex of *EcoRI* endonuclease with its DNA binding site, where an effective unwinding of about 25° occurs at the centre of the symmetrical sequence³¹. The unwinding induced by the endonuclease is important for appropriate alignment of protein-DNA interactions on either side.

The general notion that sequence-dependent aspects of DNA conformation can influence affinity of specific binding proteins has been discussed by Dickerson³². The characteristics that we see here to be necessary for a good fit, both within a half-site and between half-sites, appear to be examples of such conformational effects, but the specific 'rules' developed by Dickerson³² may not be straightforwardly applicable².

Operator recognition

The tight fit between 434 repressor and DNA could not be achieved with a random 14-bp sequence. We suppose that in a non-specific complex, the repressor would be displaced by 2–3 Å outward from the DNA axis, to accommodate non-optimal major-groove interactions, and that the peptide backbone-DNA backbone hydrogen bonds would be lost. Residues such as Arg 41 (which must curl back toward a phosphate in the specific complex) and Arg 43 (which lies in the minor groove) could extend to provide compensating electrostatic stabilization. Such non-specific complexes may be intermediates in operator binding, because proteins like 434 repressor are believed to locate their sites by initial nonspecific interaction with distant sequences, followed by diffusion along the backbone. Record *et al.*³³ have postulated a similar distinction between complexes of *lac* repressor with operator and non-operator sequences.

We do not yet have a complete description of how the 434 repressor distinguishes among its six naturally-occurring operator sites. To summarize our current picture of specific repressor-operator interactions, it is useful to distinguish three

Table 3 Amino-acid residues in α -helices

	434 repressor	λ repressor
$\alpha 1$	1–13	9–23
$\alpha 2$	17–23	33–39
$\alpha 3$	28–36	44–52
$\alpha 4$	44–53	61–69

The α backbone of λ repressor N-terminal domain⁵ was superimposed on the R1-69 backbone, optimizing coincidence of $\alpha 2$ and $\alpha 3$ (ref. 4). Assignments of helical residues were made by visual inspection of backbone conformation. Residue 15 of λ repressor closely corresponds in three-dimensional position to residue 1 of 434 repressor; 434 repressor lacks the N-terminal arm of λ repressor and the first few residues of $\alpha 1$. The first helix of λ repressor, as built by Pabo and Lewis, ends 'earlier' than $\alpha 1$ of 434 repressor (residue 23 of λ corresponds to 9 of 434). The region at the end of $\alpha 1$ was not well-defined in the original λ repressor map, however, and the actual α coincidence in $\alpha 1$ may be even greater than suggested by the assignments above. In the present 434 repressor model, residue 27 might also be assigned to helix 3, since its carbonyl appears to hydrogen bond to N31.

zones in a 7-bp half-site. (1) The first three base pairs (ACA...), invariant in all operators (see ref. 2), interact directly with Gln 28 and Gln 29 on $\alpha 3$. Changes at any one of these positions decrease repressor affinity over 150-fold (R. Wharton, personal communication), and indeed no other DNA sequence can be properly complementary to the contact surface presented by the glutamine side chains. (2) The fourth and fifth base pairs both interact with Gln 33. We therefore expect effects of altered base sequence at these positions to be correlated. One of the naturally-occurring operators (*O_R3*) differs from the others at position 4 in one half-site, and there is considerable variation among sites at position 5. (3) The sixth and seventh base pairs form no direct contact with protein, and modulation of affinity by sequence variation at these positions occurs through effects on DNA conformation. The corresponding free energy changes are modest (1–2 kcal mol⁻¹), but natural operator sites differ in binding free energy by just such small amounts. The sequences of the six sites vary widely in this zone.

The distinction between these zones is only heuristic and it is clear that a perturbation at one position will affect the fit elsewhere. For example, the entire operator half-site must be configured correctly against the repressor monomer for any of the interactions with base pairs to be meaningful. Given such a configuration, the present description appears to explain many of the observed effects of operator mutations.

We thank Marie Drott for laboratory assistance and preparation of crystals; R. Crouse for maintaining X-ray apparatus; J. Katz for artwork; A. Aggarwal, A. Mondragon, J. Moulay and G. Quigley for programs and advice; J. Koudelka and R. Wharton for discussion of unpublished results; C. Wolberger and D. C. Wiley for comments and suggestions. J. E. Anderson was a Burroughs Wellcome Fund Fellow of the Life Sciences Research Foundation. The work was supported by a grant from the NIH to S.C.H. and M.P.

Received 8 January; accepted 1 April 1987.

- Ptashne, M. *A Genetic Switch* (Cell Press, Cambridge, Massachusetts, 1986).
- Koudelka, J., Harrison, S.C. & Ptashne, M. *Nature* **326**, 886–888 (1987).
- Wharton, R. thesis, Harvard Univ. (1985).
- Anderson, J., Ptashne, M. & Harrison, S.C. *Nature* **316**, 596–601 (1985).
- Pabo, C.O. & Lewis, M. *Nature* **298**, 443–447 (1982).
- Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. *Nature* **290**, 754–758 (1981).
- McKay, D.B. & Steitz, T.A. *Nature* **290**, 744–749 (1981).
- Schevitz, R.W., Owinowski, Z., Joachimiak, A., Lawson, C.L. & Sigler, P.B. *Nature* **317**, 782–786 (1985).
- Pabo, C.O. & Sauer, R.A. *Rev. Biochem. Sci.* **53**, 293–321 (1984).
- Anderson, J., Ptashne, M. & Harrison, S.C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307–1311 (1984).
- Rossmann, M.G. & Blow, D.M. *Acta crystallogr.* **16**, 39–45 (1973).
- Bricogne, G. *Acta crystallogr.* **A130**, 395–405 (1974).
- Bricogne, G. *Acta crystallogr.* **A132**, 832–847 (1976).
- Hogle, J., Chow, M. & Filman, D. *Science* **229**, 1359 (1985).
- Jones, T.A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
- Hendrickson, W.A. & Konner, J.H. *Biomolecular Structure, Conformation, Function and Evolution* Vol. 1 (ed. Srinivasan, R.) 43–57 (Pergamon, Oxford, 1981).

- Sussmann, J., Holbrook, S.R., Church, G.M. & Kim, S.-H. *Acta crystallogr.* **33**, 800–804 (1977).
- Ten Eyck, L. *Acta crystallogr.* **A33**, 486–492 (1977).
- Hogle, J., Maeda, A. & Harrison, S.C. *J. molec. Biol.* **191**, 625–638 (1986).
- Wilson, I.A., Wiley, D.C. & Skehel, J.J. *Nature* **289**, 373–378 (1981).
- Drew, H.R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2179–2183 (1981).
- Bushman, R., Anderson, J.E., Harrison, S.C. & Ptashne, M. *Nature* **316**, 651–653 (1985).
- Hol, W.G.J., van Duijnen, P.T. & Berendsen, H.J.C. *Nature* **273**, 443–446 (1978).
- Pflugrath, J.W. & Quiocho, F.A. *Nature* **314**, 257–260 (1985).
- Dreusicke, D. & Schulz, G. *FEBS Lett.* **208**, 301–304 (1986).
- Seeman, N.C., Rosenberg, J.N. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **173**, 804–808 (1976).
- Baker, E.N. & Hubbard, R.E. *Progr. Biophys. molec. Biol.* **144**, 97–179 (1984).
- Wharton, R. & Ptashne, M. *Nature* **326**, 888–891 (1987).
- Drew, H.R. & Travers, A.A. *Cell* **37**, 491–502 (1984).
- Ryder, K., Silver, S., DeLucia, A.L., Fanning, E. & Tegtmeyer, P. *Cell* **44**, 719–725 (1986).
- McClarín, J.A. *et al. Science* **234**, 1526–1541 (1986).
- Dickerson, R.E. *J. molec. Biol.* **166**, 419–441 (1983).
- Record, M.T., deHaseth, P.L. & Lohman, T.M. *Biochemistry* **16**, 4791–4796 (1977).
- Crawford, J. thesis, Harvard Univ. (1977).
- Durbin, R. *et al. Science* **232**, 1127–1132 (1986).
- Kabsch, W. *Acta crystallogr.* **A 33**, 922 (1976).

Runaway cosmic strings

Craig J. Hogan

Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

Cosmic strings are now established as a possible mechanism for forming galaxies and clusters in the Universe (refs 1–4; H. Sato, preprint HUNS 827, Kyoto University, 1986). Much recent work has been devoted to calculating the behaviour of strings and their interaction with matter in an effort to understand better the special features of the string model. Here, I examine two effects which have been neglected in many treatments but which could, it appears, alter certain details of our picture of string-induced galaxy formation. First, loops of cosmic string behave like relativistic gyroscopes, and undergo only a very slow precession under the influence of newtonian gravitational torques. They thus impose a 'stiff' time-averaged quadrupolar gravitational potential on matter in their vicinity. Second, the gravitational radiation beamed by an oscillating loop probably has a significant component of dipole asymmetry when averaged over the loop decay time. Recoil reaction on the loop causes a coherent acceleration over time and eventually leads to a high peculiar velocity when loops decay. String loops today would then not generally be found in the centres of galaxies and clusters, but flying freely through intergalactic space.

Kibble and Turok^{5,6} showed that an ideal isolated loop oscillates in an exactly periodic orbit. I now show, in the case of one specific example and in general order-of-magnitude terms, that their result holds remarkably accurately even in the presence of gravitational torques. In particular, the mean newtonian quadrupole field of a loop varies under a torque much more slowly than that of a solid body with the same mean mass distribution. The basic reason for this behaviour is that loops have an extremely large angular momentum and act like relativistic gyroscopes. The effect of a torque is to cause a slow precession.

This effect is most clearly demonstrated in a degenerate case, that of a straight string segment in solid-body rotation in a plane. Denoting the mass per length (with $G \equiv 1$) of string at rest by μ , the spinning segment sweeps out a flat disk with radius R , mass $M = \pi\mu R$, and angular momentum $|\mathbf{J}| = \pi\mu R^2/2$. Let us apply a torque to this disk, say with an external gravitational field. By symmetry, the only components of angular momentum which can be changed are perpendicular to that of the disk. If the centre-of-mass acceleration of the string is ignored, no work is done on the segment, and to first order its mass, radius and total angular momentum $|\mathbf{J}|$ are left unchanged. Thus the segment is moving on the same degenerate trajectory as before. However, the direction of $\mathbf{J}$ is changed; the inclination of the disk changes under a torque. The rate at which the plane precesses is

$$\dot{\Theta} = |\mathbf{J}|^{-1} |\dot{\mathbf{J}}| \quad (1)$$

in a direction perpendicular both to $\mathbf{J}$ and to the applied torque.

It is reasonable to conjecture that equation (1) will give approximately the rate of rotation of newtonian quadrupole moments for less degenerate loops than the disk orbit, simply because they are usually expected to have $|\mathbf{J}|$ of the order of MR . Their angular momentum per unit mass is comparable to a maximally-rotating Kerr black hole of the same radius. Such a large angular momentum implies that loops are extremely hard to turn. For example, if one places a point mass m at a distance L (the loop length) from a loop, the rate of precession would be $\sim \mu m/LM$. For galaxy-forming loops with $\mu \approx 10^{-6}$, this is much slower than the inverse gravitational timescale $\approx \mu^{1/2}/L$ for matter orbiting the loop; it is even slow compared to the gravitational radiation decay rate $\Gamma_{\text{dec}} = \gamma\mu/L$. For typical

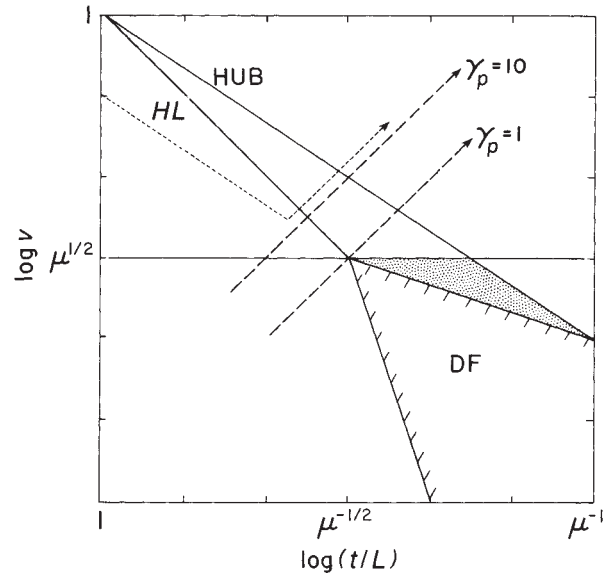


Fig. 1 Velocity as a function of time, expressed in terms of loop parameters μ and L . HL indicates the Hubble velocity across the loop length. A loop forms at $t/L \approx 1$ with a certain initial peculiar velocity v_i . It slows down by redshifting at first, parallel to HUB line. Recoil causes acceleration, as represented by $\gamma_p = 1$ and $\gamma_p = 10$ examples. The peculiar velocity of a sample loop is plotted (small-dashed lines) with $v_i \approx \mu^{1/6}$ and γ_p just greater than 10. If γ_p is less than ~ 1 and v_i is small, a loop can penetrate the hatched region where dynamical friction (DF) will slow it down. If it enters the shaded region it can accrete mass and still recoil to high velocity.

loops, Vachaspati and Vilenkin⁷ found that the numerical factor γ was usually of the order of 10^2 . Thus it takes a body considerably more massive and more dense than the loop itself to turn a loop before it decays. In any realistic situation the angular momentum of the loop dominates that of any massive body in its vicinity, such as a galaxy, and the galaxy rotates under the influence of the loop, leaving the orientation of the loop intact, rather than vice versa. The centre of mass of the loop moves normally under the gravitational influence of other bodies, but the orientation of its quadrupole field responds hardly at all to external torques.

This remarkable stability also applies to the precession of loops due to their own radiation. The centre of mass of a loop accelerates at a rate

$$\dot{\mathbf{v}}_{\text{recoil}} = \gamma_p \mathbf{n} \mu / L \quad (2)$$

where $\mathbf{n}$ is the direction of recoil, due to the beamed emission of gravitational radiation. The numerical factor γ_p is generically⁷ of the order of 10. In general, this momentum is not radiated from the centre of mass of the loop so the radiation also induces a torque. The effect is, however, more like pulling on the point of a toy gyroscope than it is like pulling on some part of a solid body. The precession rate of the figure is expected to be of the order of $\Gamma_{\text{dec}} \gamma_p / \gamma$, so the orientation of a loop generally changes more slowly than its length. This result seems to indicate that the classical precession theory would actually break down due to the mass loss from the loop before it turned appreciably.

A more adventurous conjecture based on these results (a possibility briefly noted in refs 4 and 7) is that the direction of beaming of the radiation is also coherent for a comparable time, so that the acceleration (2) is coherent in direction over a decay timescale. This would lead to a peculiar velocity $v \approx \gamma_p / \gamma$ for a loop by the time it decayed, typically of the order of 0.1. Because I have demonstrated that the global orientation of a loop is steady on a timescale Γ_{dec}^{-1} , any rapid change in beaming direction would have to arise from rapid variation in some other property of the loop oscillations. Loop oscillations can be regarded^{5,6} as waves travelling in both directions around the loop.

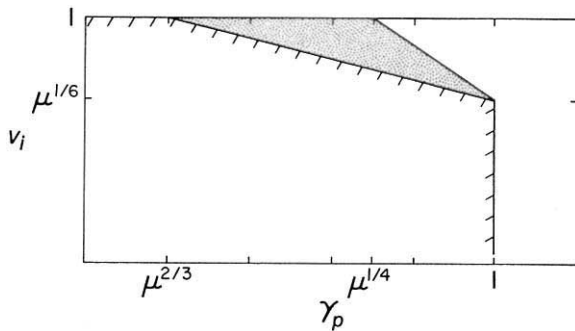


Fig. 2 Behaviour of a loop in γ_p , v_i parameter space. Hatched region represents loops which stop and accrete matter; shaded region represents loops which accrete matter and then run away; and right-hand region represents runaway loops which never accrete appreciably. This region includes the generic value $\gamma_p \approx 10$ suggested by the models of ref. 7.

High-frequency components decay rapidly, so the beaming direction is determined by the amplitude and phase of various low-order modes travelling around the loop⁷. But neither the amplitude nor the phase of these modes is expected to change by more than $\sim \gamma\mu$ per oscillation period; this would be the maximum expected 'back-reaction' on individual modes from gravitational radiation. It is possible that small changes in phase could lead to large variations in beaming direction, but if a small phase change leads to a comparable change in beaming direction, the coherence time is of the order of Γ_{dec}^{-1} . In what follows let us assume that this is the case and investigate the consequences of the resulting large secular acceleration on the motion of loops and their interaction with matter. For the following discussion to be valid, it is the only necessary for a non-negligible beamed component of radiation to have a long coherence time; short timescale fluctuations in beaming direction are allowed as long as there is a statistical long-term significant dipole anisotropy. This is a weaker condition than the conjectured stability of beaming direction, and seems very likely to hold in view of the lack of symmetry of the oscillations.

Let us then follow the centre of mass velocity $\mathbf{v}$ of a loop through its life history. Several forces act on a loop simultaneously, and it is useful to isolate one which dominates at any particular time. In addition to the recoil which accelerates the loop, it is decelerated by Hubble slowing ($\dot{\mathbf{v}} = -H\mathbf{v}$ where $H = (8\pi G\rho/3)^{1/2}$ is Hubble's constant) and by dynamical friction^{4,8,9}. If the background matter is cold with density ρ and the loop travels at least its own radius in a Hubble time ($v > HL$), we can use the Chandrasekhar friction formula

$$(\dot{v}/v)_{\text{drag}} \approx 4\pi M\rho/v^3. \quad (3)$$

Ignoring numerical factors of the order of unity, we may use the Friedmann equations to write

$$(\dot{v}/v)_{\text{drag}} \leq H(\mu HL/v^3) \quad (4)$$

where approximate equality applies for $v > HL$. For low velocities and large loops equation (4) gives only an upper bound on the friction because for the range of impact parameters contributing to the friction (that is, radii greater than L) the matter is expanding faster than v . One may still obtain a lower bound on the frictional drag rate, which is reduced by a factor of at most $\sim (v/HL)^4$ below that for a point mass in a static medium of the same density. This is because a fraction $\sim (v/HL)^2$ of the loop mass (which is effectively distributed on a two-dimensional surface⁶) near the centre of mass is still moving in an effectively cold, static medium. For $v < HL$ we then have a drag rate

$$(\dot{v}/v)_{\text{drag}} \geq H[\mu v/(HL)^3] \quad (5)$$

In addition, for $(t/L) \leq \mu^{1/2}$ once can use the results of linear

perturbation theory (see, for example, E. Bertchinger, preprint, University of California, Berkeley, 1986) to show that $(\dot{v}/v)_{\text{drag}} \leq \mu Ht/Lv$. For $(\dot{v}/v)_{\text{drag}} > H$ in this regime, v must go below the $\gamma_p = 1$ line in the left half of Fig. 1, in which case its trajectory certainly ends up dominated by the frictional effects just discussed.

Figure 1 shows the peculiar velocity of a loop as a function of time after it enters the horizon. The effects of Hubble slowing and radiation recoil re-acceleration are indicated by a sample trajectory. This example is a 'runaway' which never moves slowly enough for dynamical friction to be important. However, if a loop enters the hatched region (bounded by equations (4) and (5) with $(\dot{v}/v)_{\text{drag}} \approx H$), then dynamical friction operates at a rate greater than H and the loop will come to a stop. It is apparent that for this to occur both the recoil parameter γ_p and the initial peculiar velocity v_i must be less than certain values, otherwise the recoil effect will win and the loop ends up at high velocity.

There are some circumstances in which a loop can accrete some matter even though dynamical friction operates at a rate less than H . As long as $v \leq \mu^{1/2}$ the classical accretion radius is greater than the loop size; this is a necessary condition for any appreciable accretion, as otherwise matter cannot be accelerated to the loop's velocity in the time it takes the loop to pass. If, in addition $HL \leq \mu^{1/2}$, then the loop has a mean density greater than the background matter and we can ignore the expansion of the background. In this case we estimate the mass accretion rate using the Bondi formula

$$\dot{M}/M \approx H(\mu HL/v^3) \quad (6)$$

In Fig. 1, the shaded region is a regime where loops are accreting some matter, albeit at a rate $< H$, but where they are still liable to be accelerated by recoil rather than coming to a stop, as the Chandrasekhar rate (equation (4)) is also less than H . These loops accrete less than their own mass in ordinary matter, then (if γ_p is large enough) accelerate to high velocity. A loop will carry along its accreted matter to high velocity as long as the 'internal' gravitational acceleration $\sim \mu^{1/2}/L$ exceeds the recoil acceleration $\sim \mu\gamma_p/L$.

Figure 2 summarizes the general behaviour of a loop as a function of its recoil parameters and its initial peculiar velocity. In this brief discussion various effects have been ignored, such as the relative alignment of the various vector quantities (which is important in those rare instances where $\mathbf{n}$ is nearly antiparallel to the initial peculiar velocity), the behaviour of hot or gaseous background matter (which reduces the frictional drag), the effects of a clumpy matter distribution, and various numerical factors. But it is clear that if asymmetrical beaming is generally as strong as that found by Vachaspati and Vilenkin⁷ and as coherent as conjectured here, then only a small fraction of loops end their lives stationary. It is worth noting that although the end of a loop's life-history is considerably altered by the runaway phenomenon, the seeding of mass-density fluctuations, which often occurs early in the life history of a loop, is not necessarily strongly affected. On the other hand if decaying loops are responsible for jets in quasars¹⁰, it is puzzling that they should often be found in the centres of galaxies. Most small loops should be found flying through intergalactic space, and those which are captured in galaxies and clusters should be found away from the centre of the potential well, where the gravitational force of the galaxy balances the recoil force. If the shape of the potential does not admit such an equilibrium, then even captured loops can sometimes escape.

There are two possible alternatives to the runaway model. (1) Strongly beaming loop configurations may not be generic; it is possible, for example, that features in the oscillations responsible for the beaming are damped in much less than a decay time and most loops quickly occupy a region of parameter space with $\gamma_p < 1$. If this happens early enough, beaming can be ignored entirely. (2) It is possible that beaming remains strong

but its direction changes rapidly and contains little long-term statistical dipole anisotropy. Neither of these alternatives seems likely to occur; however, detailed study of gravitational back-reaction is necessary to establish with certainty that the runaway effects described here are indeed significant.

I thank S. D. M. White and A. Vilenkin for useful discussions. This work was supported by NASA and by the theoretical astrophysics program of the University of Arizona.

Received 5 February; accepted 24 February 1987.

1. Vilenkin, A. *Phys. Rep.* **121**, 263–315 (1985); Vilenkin, A. in *Inner Space/Outer Space* (eds Kolb, E. W. et al.) 269 (University of Chicago Press, 1986).
2. Turok, N. *Phys. Rev. Lett.* **55**, 1801 (1985).
3. Turok, N. & Brandenberger, R. H. *Phys. Rev. D* **33**, 2175 (1986).
4. Hogan, C. J. & Rees, M. J. *Nature* **311**, 109 (1984).
5. Kibble, T. W. B. & Turok, N. *Phys. Lett.* **116B**, 141 (1982).
6. Turok, N. *Nucl. Phys. B* **242**, 520 (1984).
7. Vachaspati, T. & Vilenkin, A. *Phys. Rev. D* **31**, 3052 (1985).
8. Silk, J. & Vilenkin, A. *Phys. Rev. Lett.* **53**, 1700 (1984).
9. Hogan, C. J. *Phys. Lett.* **143B**, 87 (1984).
10. Vilenkin, A. & Field, G. *Nature* (in the press).

Flux quantization in a high- T_c superconductor

C. E. Gough*, M. S. Colclough*, E. M. Forgan*,
R. G. Jordan*, M. Keene*, C. M. Muirhead*,
A. I. M. Rae*, N. Thomas*, J. S. Abell† & S. Sutton†

* Department of Physics, University of Birmingham,
Birmingham B15 2TT, UK

† Department of Metallurgy and Materials, University of
Birmingham, Birmingham B15 2TT, UK

We have observed the quantization of magnetic flux in a high- T_c yttrium-based ceramic superconductor¹ and obtain a value for the flux quantum in this material. The value of the flux quantum, $h/2e$ where h is Planck's constant and e is the electron charge, implies that the charge carriers of superconductivity are electron pairs.

The measurements were made on a ring (outer-diameter 10.0 mm, inner-diameter 4.5 mm and thickness 4.0 mm) formed into this shape by sintering $Y_{1.2}Ba_{0.8}CuO_4$ prepared from powders of $BaCO_3$, Y_2O_3 and CuO heated in air at 950 °C. The sample had a broad superconducting transition between 50 and 85 K.

The ring was immersed in liquid helium at 4.2 K inside a superconductivity shield. The flux inside the ring was monitored by a radio frequency superconducting quantum interference device (r.f.-SQUID) magnetometer weakly coupled to the ring via a flux transformer and flux was applied to the ring by a solenoid passing through it. As the current in the solenoid was changed, supercurrents were induced in the ring until some critical current was exceeded when flux passed in or out of it. A typical magnetization curve obtained in this way is shown in Fig. 1. Such behaviour is characteristic of measurements made on a superconducting circuit containing a weak link².

The use of the long solenoid (diameter 1.86 mm with a turn spacing of 0.142 mm) bent into a nearly complete loop passing through the centre of the ring allows us to make an accurate determination of the applied flux. Along the sloping sections of the magnetization curve the total flux remains constant and the induced magnetometer signal is therefore determined entirely by the flux in the solenoid. We note that there is a slight overall slope to the magnetization curve due to a small amount of direct coupling between the solenoid and the flux transformer. After allowing for this we obtain an accurate calibration of the magnetometer signal in terms of the flux produced by the circulating supercurrent whatever its actual path through the ring. In this

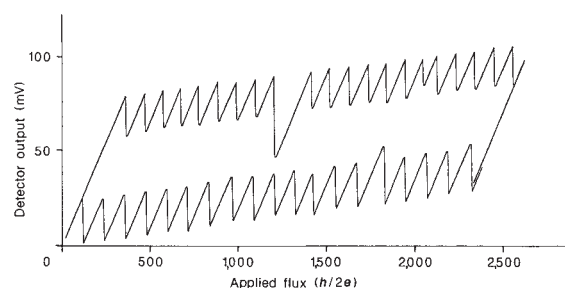


Fig. 1 Typical magnetization curve showing the flux output of the r.f.-SQUID magnetometer as the flux applied through the long solenoid is swept through a single cycle.

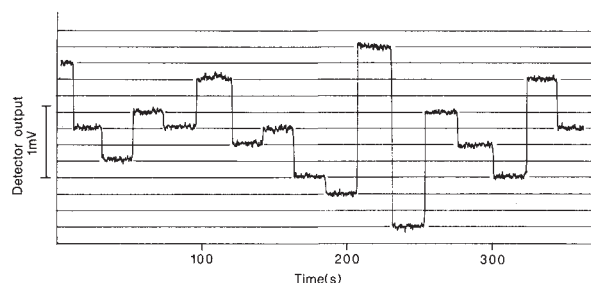


Fig. 2 Output of the r.f.-SQUID magnetometer showing small integral numbers of flux quanta jumping in and out of the ring.

way we deduce that the magnitude of the flux jumps shown in Fig. 1 is typically 100 ($h/2e$).

To obtain a value for the flux quantum, we set the solenoid current to zero and periodically expose the ring to a local source of electromagnetic noise causing the ring to make small jumps between quantized flux-states. This is illustrated in Fig. 2, in which a number of equally-spaced lines have been superimposed to emphasize the quantum nature of the flux transitions. Using these results and the earlier calibration we obtain a value for the flux quantum of 0.97 ± 0.04 ($h/2e$) thus demonstrating that superconductivity in this high- T_c material involves the pairing of electrons as in conventional BCS (Bardeen-Cooper-Schrieffer) superconductivity. The observed flux jumps may be between the outside of the ring and the hole, or may be between the hole and the bulk of the material. We are not able to distinguish between these two possibilities at present.

We have performed experiments similar to those shown in Fig. 2 in which we observed no detectable change in the magnetometer output for 1,000 s. On this basis the resistance of our sample is less than 10^{-13} ohms.

The above measurements are of interest not only because they demonstrate the existence of coherent superconducting states in the newly discovered class of ceramic superconductors but also because they suggest that SQUID-type devices could be made from simple fabricated shapes using the intrinsic properties of the weakly superconducting ceramic material to provide its own weak link. Indeed, we have already used our ring to observe noisy r.f.-SQUID characteristics at 20 MHz typical of a weak-link loop with too large a critical current. We are currently extending all these measurements to other rings and to higher temperatures to investigate the compositional and temperature dependence of the magnetic and SQUID behaviour.

We thank G. R. Walsh for technical assistance and Professor W. F. Vinen for support and encouragement.

Received 8 April; accepted 10 April 1987.

1. Wu, M. K. et al. *Phys. Rev. Lett.* **58**, 908–911 (1987).
2. Zimmerman, J. E. & Silver, A. H. *Phys. Rev.* **157**, 317–341 (1967).

Identification of the phase responsible for high-temperature superconductivity in Y-Ba-Cu oxides

C. N. R. Rao*, P. Ganguly*, A. K. Raychaudhuri†, R. A. Mohan Ram* & K. Sreedhar*

* Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India

† Department of Physics, Indian Institute of Science, Bangalore 560 012, India

The discovery of La-Ba(Sr)-Cu oxides of K_2NiF_4 structure exhibiting superconductivity in the 30–40 K region was reported early this year¹. The recent announcement of superconductivity above the temperature of liquid nitrogen in Y-Ba-Cu oxides² has belied all expectations. Unfortunately, all the Y-Ba-Cu oxide compositions exhibiting this remarkable feature were biphasic or multiphasic. We have studied various compositions of the $Y_{3-x}Ba_{3+x}Cu_6O_{14}$ and now report the isolation of the pure oxide phase responsible for high T_c superconductivity. This oxide has the composition $Y_{0.33}Ba_{0.67}CuO_{2.33 \pm 0.05}$, with an oxygen defect perovskite structure. The pure oxide shows the onset of superconductivity at 120 K, attaining zero resistance at 87 K. It also exhibits the highest Meissner effect of all the high- T_c oxide superconductors.

High-temperature superconductivity in the oxide composition $Y_{1.2}Ba_{0.8}CuO_4$ was reported a few weeks ago by Wu *et al.*²; the material attained zero resistance at 83 K. This oxide composition which was made by analogy with the oxides of K_2NiF_4 structure^{1,3} was actually biphasic consisting of a green and a black oxide. We have been investigating superconductivity in Y-Ba-Cu oxides analogous to the defect perovskite $La_3Ba_3Cu_6O_{14}$ (Fig. 1) because Y_2CuO_4 does not form in the K_2NiF_4 structure. The Ba ions in this oxide are in the 10-coordinated A_1 sites and the La ions in the 8-coordinated A_2 sites. The copper ions have three possible coordinations, but they show some features of the K_2NiF_4 structure with essentially a square-planar coordination and no long-range Cu-O-Cu interaction along the c -axis. Studies on various compositions of the general formula $Y_{3-x}Ba_{3+x}Cu_6O_{14 \pm \delta}$ revealed them to be superconducting with zero resistance in the 80–90 K range⁴, but they were all biphasic.

To establish the identity of the oxide responsible for the high-temperature superconductivity, we have carried out an X-ray diffraction study of several members of

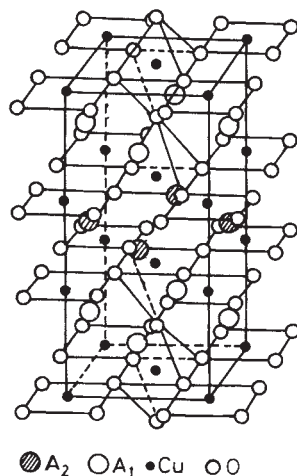


Fig. 1 Structure of $La_3Ba_3Cu_6O_{14 \pm \delta}$ with La in both A_1 and A_2 sites (after Er-Rakho *et al.*⁸). The yttrium analogue cannot be formed because Y cannot go to the A_1 sites. However, in $Y_2Ba_4Cu_6O_{14 \pm \delta}$, Y is only in A_2 sites and Ba in A_1 sites.

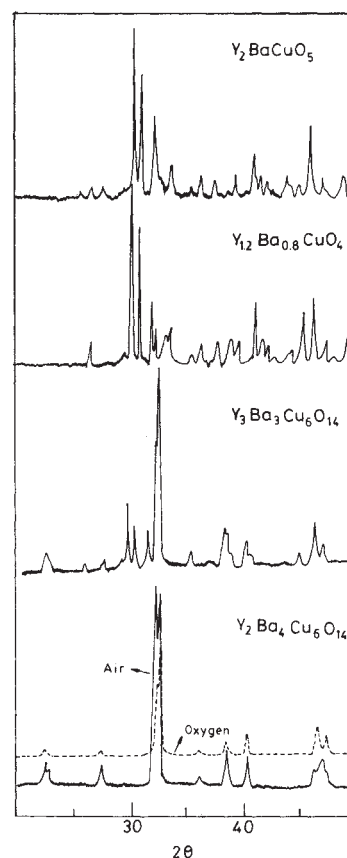


Fig. 2 X-ray diffraction patterns of Y_2BaCuO_5 , $Y_{1.2}Ba_{0.8}CuO_4$ (prepared in air at ~1,300 K, 24 h), $Y_3Ba_3Cu_6O_{14 \pm \delta}$ (prepared in air at ~1,200 K, 12 h), $Y_2Ba_4Cu_6O_{14 \pm \delta}$ (prepared in air at 1,200 K, 12 h) and $Y_2Ba_4Cu_6O_{14 \pm \delta}$ (annealed in oxygen at 1,100 K for 6 h after preparing it at 1,200 K in air).

$Y_{3-x}Ba_{3+x}Cu_6O_{14 \pm \delta}$, $Y_{1.2}Ba_{0.8}CuO_4$ and the green-coloured insulator Y_2BaCuO_5 . The study shows that the $Y_{1.2}Ba_{0.8}CuO_4$ composition determined by Wu *et al.*² consists of a high proportion of Y_2BaCuO_5 and a small proportion of a perovskite, whereas $Y_{3-x}Ba_{3+x}Cu_6O_{14 \pm \delta}$ consists mainly of the perovskite phase (Fig. 2). Based on the relative intensities of the X-ray lines, we derive the composition of the perovskite phase to be $Y_2Ba_4Cu_6O_{14 \pm \delta}$, or $Y_{0.33}Ba_{0.67}CuO_{2.33 \pm 0.05}$. The structure of this perovskite can be understood in terms of that of $La_3Ba_3Cu_6O_{14}$. Because the yttrium ions cannot be on the A_1 sites, the observed ratio of Y and Ba ions is exactly the same as the ratio of the available A_2 and A_1 sites (1:2).

We have prepared $Y_{0.33}Ba_{0.67}CuO_{2.33 \pm 0.05}$ by heating a mixture of the component oxides at 1,200 K in air. The X-ray diffraction pattern (Fig. 2) clearly establishes it to be a perovskite. The pure perovskite, however, seems to be more distorted than the perovskite phase in the $Y_{3-x}Ba_{3+x}Cu_6O_{14 \pm \delta}$ compositions. When $Y_{0.33}Ba_{0.67}CuO_{2.33 \pm 0.05}$ was annealed in a stream of oxygen at 1,100 K, the perovskite showed less distortion as indicated in Fig. 2. It appears that the structure of the perovskite is very sensitive to oxygen stoichiometry. We are now carrying out a refinement of the structure of the perovskite to establish the atomic positions.

The room-temperature resistivity of $Y_{0.33}Ba_{0.67}CuO_{2.33 \pm 0.05}$ prepared in air at 1,200 K was $\sim 4 \times 10^{-2}$ ohm cm and the material did not show high T_c superconductivity. But the sample treated in oxygen had a lower room temperature resistivity ($\sim 4 \times 10^{-3}$ ohm cm) and showed the onset of superconductivity at 120 K and zero resistance at 87 K (Fig. 3). Heating in a stream of oxygen seems to bring the oxygen stoichiometry and hence the Cu^{2+}/Cu^{3+} ratio (with nearly 33% Cu^{3+}) to that required

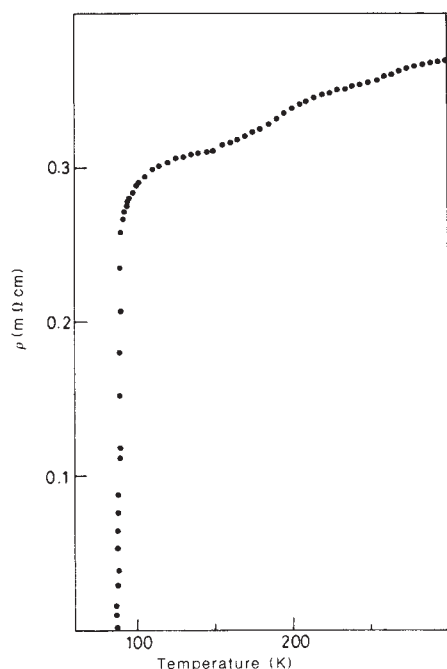


Fig. 3 Resistivity data of $\text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{14\pm\delta}$ (prepared in air at 1,200 K and then annealed in oxygen at 1,100 K, 6 h) showing the onset of superconductivity at 120 K and zero resistance at 87 K. Zero resistance is found at 91 K on prolonged heating of oxygen.

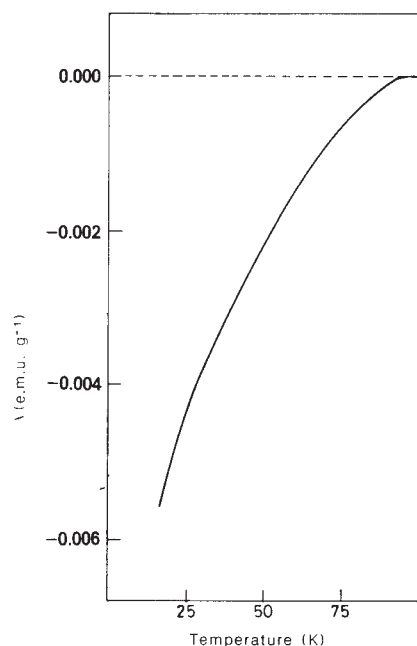


Fig. 4 Magnetic susceptibility data of superconducting $\text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{14\pm\delta}$ (measured in ~ 500 G field). The a.c. magnetic susceptibility data and data measurements show 100% Meissner effect in the sample.

for high-temperature superconductivity to occur. It is noteworthy that the room temperature resistivity of this superconducting sample is on the borderline of the metal-nonmetal transition^{5,6}. It seems essential for the resistivity to be in this region for the manifestation of superconductivity in these materials. We had earlier noticed this behaviour in the superconducting oxides of the type $\text{La}_{2-x}\text{Ba}_x(\text{Sr}_x)\text{CuO}_4$ possessing the K_2NiF_4 structure^{3,7}. However, the Cu^{3+} content in $\text{Y}_{0.33}\text{Ba}_{0.67}\text{CuO}_{2.33\pm\delta}$ is considerably higher ($\geq 33\%$) than in the oxides of K_2NiF_4 structure ($\sim 15\%$), the superconducting transition temperature is also much lower in the latter.

Magnetic susceptibility measurements on $\text{Y}_{0.33}\text{Ba}_{0.67}\text{CuO}_{2.33\pm\delta}$ are shown in Fig. 4. The d.c. susceptibility is strongly field-dependent even at very low temperatures. The susceptibility at 15 K is around 50% of the full Meissner value. The temperature dependence of the susceptibility is reminiscent of a Type II superconductor. The Meissner effect found by us in this study is probably the highest reported in oxide superconductors.

Electrical and magnetic properties of $\text{Y}_{1.9}\text{Ba}_{4.0}\text{Cu}_6\text{O}_{14\pm\delta}$ and $\text{Y}_{2.1}\text{Ba}_{3.9}\text{Cu}_6\text{O}_{14\pm\delta}$ were measured to examine the behaviour of compositions on either side of $\text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{14\pm\delta}$. The $\text{Y}_{1.9}$ sample shows zero resistance at 84 K while the $\text{Y}_{2.1}$ sample shows it at 82 K and the superconducting transition in both the samples was not as sharp as in $\text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{14\pm\delta}$; the Meissner effect was also smaller in the first two samples.

Our preliminary low-temperature electron diffraction studies of $\text{Y}_{0.33}\text{Ba}_{0.67}\text{CuO}_{2.33\pm\delta}$ have shown no evidence for any significant change in the structure. We are now carrying out nuclear magnetic resonance optical and other spectroscopic experiments as well as neutron diffraction studies on this perovskite.

We thank the Department of Science and Technology, Government of India, and the University Grants Commission, for support.

Received 26 March; accepted 9 April 1987.

1. Strongin, M., Welch, D. O. & Davenport, J. W. *Nature* **325**, 664–665 (1987).
2. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908–911 (1987).
3. Rao, C. N. R. & Ganguly, P. *Curr. Sci.* **56**, 47–49 (1987).
4. Ganguly, P., Raychaudhuri, A. K., Sreedhar, K. & Rao, C. N. R. *Pramana-J. Phys.* **27**, 229–232 (1987).
5. Mott, N. F. *Metal-Insulator Transition* (Taylor & Francis, London, 1974).
6. Rao, C. N. R. & Ganguly, P. in *The Metallic and the Nonmetallic States of Matter* (eds Edwards, P. P. & Rao, C. N. R.) (Taylor & Francis, London, 1985).
7. Ganguly, P., Mohan Ram, R. A., Sreedhar, K. & Rao, C. N. R. *Solid St. Commun.* (in the press).
8. Er-Rakho, L., Michel, C., Provost, J. & Raveau B. *J. Solid St. Chem.* **37**, 151–157 (1981).

High superconducting transition temperatures in sputter-deposited YBaCuO thin films

R. E. Somekh, M. G. Blamire, Z. H. Barber, K. Butler, J. H. James, G. W. Morris, E. J. Tomlinson, A. P. Schwarzenberger, W. M. Stobbs & J. E. Evetts

Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

The recent discovery of a remarkable new family of high critical temperature superconducting oxides has excited extraordinary interest and research activity worldwide^{1–6}. The prospect of superconducting devices and machines operating in liquid nitrogen and at higher temperatures is seen as an unprecedented technological development. There are potential applications in microelectronics and computers, energy storage and transmission, transport, high technology engineering, sensors and instrumentation. For many applications it is essential to prepare these materials in the form of thin films on suitable substrates. Currently, most of the research effort and all of the available published information relates to the fabrication of bulk samples from powders. At Cambridge our primary objective has been to achieve the deposition of high-quality thin films. We have taken this course because, (1) deposition from a multi-target sputter system gives great flexibility and enables us to cast a wide parameter net in composition and temperature to ensure early identification of key superconducting phases, (2) the deposition and processing of $0.5\text{ }\mu\text{m}$ films can be carried out very rapidly with one or possibly more system cycles per day and (3) measurements on thin films provide quantitative information on basic superconducting parameters and are likely to lead to the

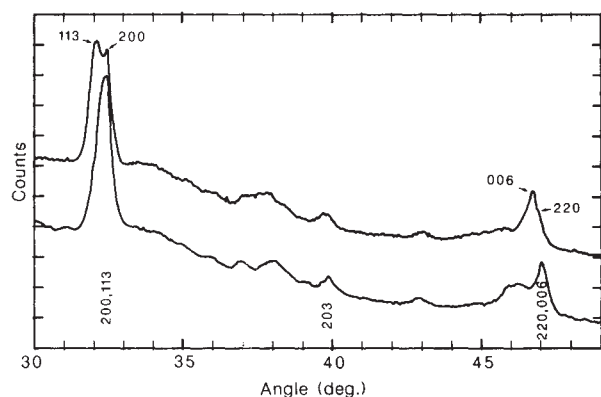


Fig. 1 X-ray diffractometer traces (copper $K\alpha$ radiation) for the thermally cycled sample shown in Table 1 following a 5 min heat treatment at 560 °C in oxygen (lower trace) and 5 min heat treatment in argon at 560 °C (upper trace). Sample composition was $Y_{0.16}Ba_{0.29}Cu_{0.55}O_y$. The peaks show a systematic angle shift in 2θ indicating a change in the lattice constants associated with the oxygen removal. In addition there is increased splitting of the two strongest lines. The upper trace is vertically offset from the lower trace by two divisions (upwards).

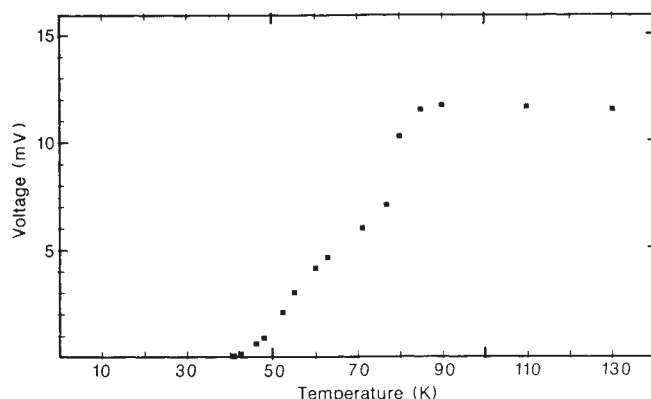


Fig. 2 Four terminal resistive transition for a sample with mean composition $Y_{0.17}Ba_{0.34}Cu_{0.5}O_y$ sample. The contact separation was 2 mm on a 3 mm wide by 500 nm thick film measured at 10 pA.

first practical applications of these remarkable materials. Here we report preliminary results on the deposition of relatively homogeneous films with transition temperature onsets up to 99 K. First results are reported on film deposition conditions, film composition and structure, and on the effect of post-deposition anneals on superconducting properties.

The samples were deposited onto R-plane sapphire substrates placed on a heater at 1,050 °C in a four target UHV d.c.-magnetron getter-sputter deposition system. The 0.5–0.6 μm thick films were deposited at a target-substrate distance of 120 mm and a rate of 0.2 to 0.3 nm s^{-1} . In any one deposition run a range of composition was produced, varying along the length of the heater. The base pressure was 10^{-7} Pa, but with the heater at temperature 1,050 °C this increased to between 2 and 5×10^{-7} Pa (see ref. 7). The outgassing rate of the system remained below 10^{-7} Pa s^{-1} . Oxide films were formed using two methods. In the first method oxygen was continuously introduced at 0.08 Pa (on 4 Pa argon) throughout the deposition. The second method was to deposit a 20–30 nm thick metal film in argon and then introduce oxygen. The oxidation reaction was observed as a fall in oxygen pressure over a period of 5–10 min. The sputtering targets, as well as the samples, were oxidized during this time and took several minutes to reach a steady state during subsequent deposition of the next alloy layer.

The published results on high critical temperature (T_c) oxides are characterized by extensive sample inhomogeneity and the proliferation of different layer structure phases and other oxides. It is therefore important to establish the structure and composition of the various superconducting phases. After removal from the deposition system the composition of representative samples was determined using energy dispersive X-ray (EDX) Analysis on an ISI 100A scanning electron microscope. In addition, a preliminary structural investigation was carried out using X-ray diffraction and transmission electron microscopy (TEM). The composition varied by up to 10% along the length of a single $12.5 \times 3 \text{ mm}^2$ substrate. A TEM investigation of one film with an onset T_c of 75 K indicated the presence of at least two majority phases and up to six minority phases including a hexagonal layer structure. The grain size was of the order of the film thickness in the high- T_c region of the film. But, X-ray diffractometer traces from ten samples consistently show the same peaks and the main lines are similar to those observed for high- T_c bulk samples by S. B. Qadri *et al.* (personal communication) at the Naval Research Laboratory, Washington their indexing of the trace is based on the analysis of a possible structure for

which detailed data have been given by Er-Rakho *et al.*⁸. The lower curve in Fig. 1 shows a representative X-ray trace, which also shows some minor peaks associated with minority phases.

Analysis of the variation of the resistive transition with composition leads us to conclude that the high- T_c phase is $YBa_2Cu_3O_7$. This is inferred from the observation that the highest T_c in any given run always has a composition with Y:Ba ratio 1:2 and Cu in the range 55–60%. The structure and characteristics of this phase have been reported by Michel and Raveau⁹. The compositions surveyed in our study ranged from $Y_{0.3}Ba_{0.1}Cu_xO_y$ to $Y_{0.1}Ba_{0.3}Cu_xO_y$, x lying between 0.4 and 0.85.

In spite of this inhomogeneity, there is good evidence that a substantial proportion of the material is superconducting, and behaves in a manner consistent with that of the perovskite layer structures described elsewhere⁹. In particular, the superconducting transition shown in Fig. 2 shows at least two superconducting transitions associated with different phases during the fall in resistance. The breadth of this transition is at least partly due to the two-dimensional grain structure of the film; in two dimensions at least 50% of randomly distributed material must be superconducting for a complete superconducting percolation path to exist, compared with only 17% for three-dimensional distribution of grains in a bulk sample¹⁰.

One significant characteristic of these compounds is the reversible uptake of oxygen and associated changes in electrical conductivity^{9,11}. Table 1 summarizes a series of heat treatments

Table 1 Two-terminal resistance changes following various heat treatments

No.	Treatment	Room temperature resistance (Ω)
0	As deposited	1,400
1	5 min Ar 500 °C	129×10^3
2	5 min O ₂ 500 °C	560
3	5 min O ₂ 500 °C	590
4	5 min Ar 560 °C	560×10^3
5	5 min O ₂ 560 °C	590

carried out on one slide. This shows the expected reversibility and the dramatic decrease in conductivity associated with oxygen depletion. The superconducting transition was measured twice, before and after this series of treatments. Before heat treatment the sample had a T_c onset of 20 K and was only partially superconducting at 4.2 K, after the final oxygen anneal it had a T_c onset of 75 K and was found to be fully superconducting at 15 K. Annealing and reversibility experiments were essentially complete after 5 min at 500 °C, indicating the potential

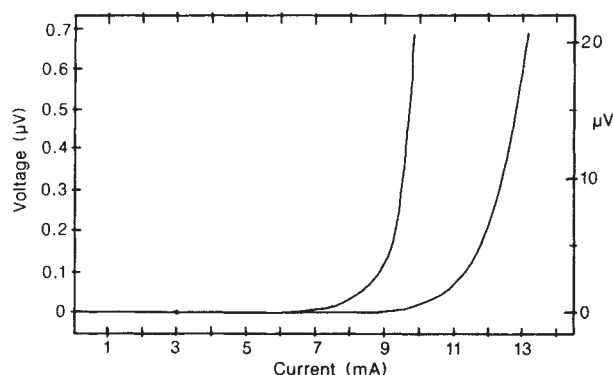


Fig. 3 Voltage against current curve at 4.2 K for sample described in Fig. 1. The curve was recorded at two different voltage sensitivities: 1 μV full scale deflection (left trace) and 30 μV full scale deflection (right curve). It was fully superconducting below 7 mA.

value of thin film samples in studying these compounds. To achieve the same result in bulk samples an anneal of the order of 24 h is required¹¹. The results allow the very high diffusion rate for oxygen intercalation at 500 °C to be estimated as about $10^{-15} \text{ m}^2 \text{ s}^{-1}$. A comparison of the X-ray diffractometer traces in the oxygen-rich (superconducting) and oxygen-poor (semi-conducting) states unexpectedly showed that most of the peaks are shifted to higher d -spacings on removal of oxygen (Fig. 1). Preliminary analysis based on the data of Qadri and co-workers suggests that the c -spacing expands marginally more than the a -spacing resulting in a further splitting of the $\{113, 200\}$ and the $\{006, 220\}$ lines on removal of oxygen.

A further demonstration of the unusual behaviour of these materials is provided by a one minute immersion in water. This completely, and apparently irreversibly, transformed the black conducting film into a largely transparent and insulating layer. This behaviour, together with the reversible oxygen uptake, indicate the presence of extremely active defect sites within the crystal lattice.

Critical temperature measurements were performed on numerous samples, both as deposited and following heat treatment in oxygen. Figure 2 shows the sharpest transition so far obtained, the material between the voltage contacts becoming fully superconducting at 36 K. At this temperature the critical current along the slide was approximately 3 mA. If the whole width of the slide were assumed to be uniformly superconducting this implies a critical current density of 60 A cm^{-2} . This is very much a lower bound estimate in view of the inhomogeneity of the two-dimensional film. Figure 3 shows the voltage-current curve for a sample at 4.2 K plotted at two different voltage sensitivities. This indicates that a complete superconducting state exists up to 7 mA with a sharp onset of flux flow at a well-defined critical current. The lower bound placed on the resistivity by these measurements is $\sim 10^{-12} \Omega \text{ m}$. The resistivity of the normal material at room temperature is $\sim 5 \times 10^{-5} \Omega \text{ m}$. The lower sensitivity curve shows that the plot continues to curve upwards even at high voltages: this is again indicative of inhomogeneity. We have observed complete superconductivity in materials with measured Y:Ba ratios which lie between 1:1.6 and 1:3. The sharpest transitions have been obtained from samples with the composition $\text{Y}_{0.17}\text{Ba}_{0.34}\text{Cu}_{0.50}\text{O}_y$.

The present results have demonstrated the practical deposition of high T_c -oxide films and give a strong indication of the composition of the high T_c phase. Future work will be aimed at increasing film homogeneity and perfection with view to carrying out basic superconductivity measurements, investigating the effect of additives and fabricating tunnel junctions and devices. We would like to thank Dr J. M. Lumley for helpful discussions during the course of this programme.

Received 1 April; accepted 9 April 1987.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189-193 (1986).
2. Uchida, S., Takagi, H., Kitazawa, K. & Tanaka, S. *Jap. J. appl. Phys.* **26**, 21-24 (1987).
3. Chu, C. W. *et al. Phys. Rev. Lett.* **58**, 405-407 (1987).
4. Cava, R. J., van Dover, R. B., Batlogg, B. & Rietman, E. A. *Phys. Rev. Lett.* **58**, 408-410 (1987).
5. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908-910 (1987).
6. Zhao, Z. *et al. Kexue Tongbao* (in the press).
7. Somekh, R. E. & Baxter, C. S. *J. Cryst. Growth* **76**, 119-125 (1986).
8. Er-Rakho, L., Michel, C., Provost, J. & Raveau, B. *J. Solid St. Chem.* **37**, 151-157 (1981).
9. Michel, C. & Raveau, B. *Rev. Chim. miner.* **21**, 407-425 (1984).
10. Evetts, J. E. *IEEE Trans. Magn.* **19**, 1109-1119 (1983).
11. Tarascon, J. M., Greene, L. H., McKinnon, W. R., Hull, G. W. & Geballe, T. H. *Science* **235**, 1373-1376 (1987).

New layer structures in the La-Cu-O system

A. H. Davies & R. J. D. Tilley

Institute of Materials, University College, Newport Road, Cardiff CF2 1TA, UK

Recently it was reported that samples with overall compositions given by $\text{La}_{1-x}\text{Ba}_x\text{CuO}_{3-y}$ were superconducting at the unprecedentedly high temperature of approximately 35 K¹. A following report in January 1987², indicated a superconducting temperature of 52.5 K for oxides in the same quaternary system. In this second report the phase which was supposed to be superconducting was a non-stoichiometric oxide of composition $(\text{La}_{0.9}\text{Ba}_{0.1})_2\text{CuO}_{4-y}$. A number of other reports, for example, ref. 3, suggest that compounds exhibiting superconducting temperatures of 100 K or higher are to be expected soon. The exact nature of the superconducting phase or phases in these materials is unknown. Until crystallographically well-characterized samples are used in conductivity measurements, the possibility of constructing a theory which accurately explains the phenomenon of superconductivity in these materials will be difficult, if not impossible. Moreover, the search for new materials with enhanced properties, and the preparation of materials in a form suitable for device use will have to be approached empirically. To provide these data, a systematic investigation of phase relations in the systems of interest has been started. Here we present new data on the parent La-Cu-O system which give evidence of the existence of new layer structures of relevance to the superconducting systems described above.

The La_2CuO_4 phase has a distorted K_2NiF_4 structure with an orthorhombic unit cell, with dimensions $a = 0.536 \text{ nm}$, $b = 0.541 \text{ nm}$, $c = 1.317 \text{ nm}$, $Z = 4$ (see ref. 4). It is presumed that the Jahn-Teller effect at copper ion sites is responsible for the distortion. The K_2NiF_4 structure type can be thought of as the first member of a homologous series of phases of composition

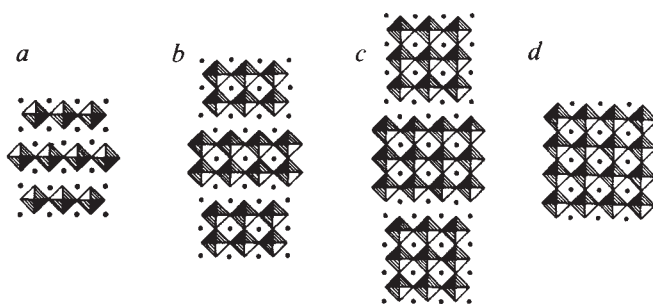


Fig. 1 Idealized representations of the structures of the series $\text{La}_{n+1}\text{Cu}_n\text{O}_{3n+1}$. a, La_2CuO_4 ; b, $\text{La}_3\text{Cu}_2\text{O}_7$; c, $\text{La}_4\text{Cu}_3\text{O}_{10}$; d, LaCuO_3 . The shaded squares represent CuO_6 octahedron and the filled circles the La atoms. In a-c the tetragonal cell is projected down $[010]$ and the c -axis is vertical.

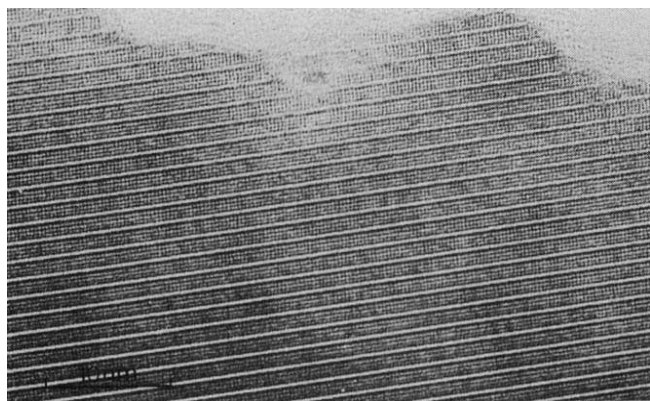


Fig. 2 Electron micrograph of a well-ordered crystal fragment of $\text{La}_4\text{Cu}_3\text{O}_{10}$. The white stripes delineate the lamellae of perovskite type and are 1.42 nm apart.

$\text{A}_{n+1}\text{B}_n\text{O}_{3n+1}$, which consists of slabs of the perovskite structure type separated by layers of the NaCl structure type (see for example, ref. 5). The first three members of the series are shown in Fig. 1. Although a series of phases belonging to this family are known to occur in the related systems Sr-Ti-O (ref. 5) and La-Ni-O (ref. 6) it is not known whether a homologous series of K_2NiF_4 -related phases exists in the La-Cu-O system.

To synthesize possible members of the series, mixtures of La_2O_3 and CuO corresponding to compositions between La_2CuO_4 and $\text{La}_2\text{Cu}_2\text{O}_5$, were ground in an agate mortar, pelleted and heated in air at 950 °C and 900 °C for 96 hours. After heating, the pellets were quenched into liquid nitrogen. Samples were examined by transmission electron microscopy using a JEM 200CX microscope fitted with a top entry goniometer stage capable of $\pm 10^\circ$ tilt and a resolution of 0.14 nm. Electron microscope samples were prepared by grinding small quantities of material in an agate mortar under *n*-butanol and allowing a drop of the resultant suspension to dry on a porous carbon support film. Samples were also examined by X-ray powder diffraction using a Hägg-Guinier focusing camera and strictly monochromatic $\text{CuK}\alpha_1$ radiation.

A comparison of the X-ray films with those of the isostructural Sr-Ti-O phases showed that long unit cell phases occurred in the 900 °C preparations which resembled members of the $\text{A}_{n+1}\text{B}_n\text{O}_{3n+1}$ structural series. The powder data will be published elsewhere. In contrast to this, these phases did not occur at 950 °C and all samples contained only La_2CuO_4 and CuO. It thus appears that the homologous series is only stable up to a temperature of $\sim 900^\circ\text{C}$ and that preparation conditions are of importance if these structures are to be synthesized.

To confirm the identity of the phases electron microscopy was used. To simplify interpretation of image contrast in these initial experiments, only a limited number of {001} systematic diffracted beams was allowed to contribute to the images. This procedure reveals the perovskite lamellae clearly. The thicknesses of these lamellae, *t*, were calculated theoretically from the structural characteristics of the phases and the baseline data given by the crystallographic parameters of La_2CuO_4 . They were found to be, for $n=1$, $t=0.66$ nm; $n=2$, $t=1.04$ nm; $n=3$, $t=1.42$ nm; $n=4$, $t=1.80$ nm and so on.

Micrographs showed that all the crystal fragments examined were composed of parallel stackings of lamellae of thicknesses corresponding to that of the perovskite slabs given above. It is proved therefore that a homologous series does exist in the system at temperatures up to $\sim 900^\circ\text{C}$. The phase La_2CuO_4 itself was usually well-ordered and had a structure similar to that represented in the idealized diagram of Fig. 1a. The other new

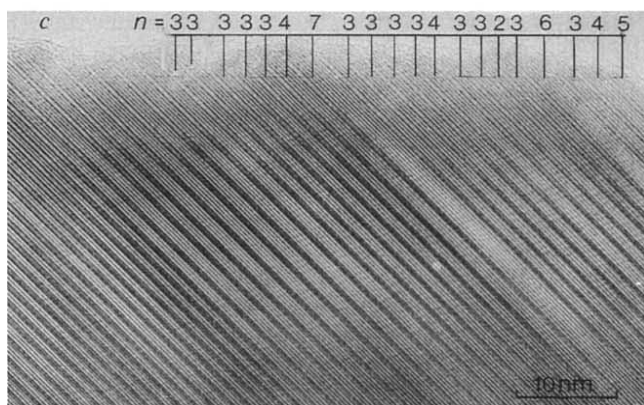
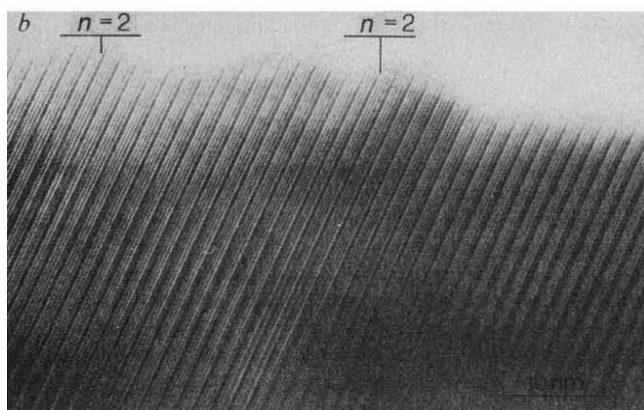
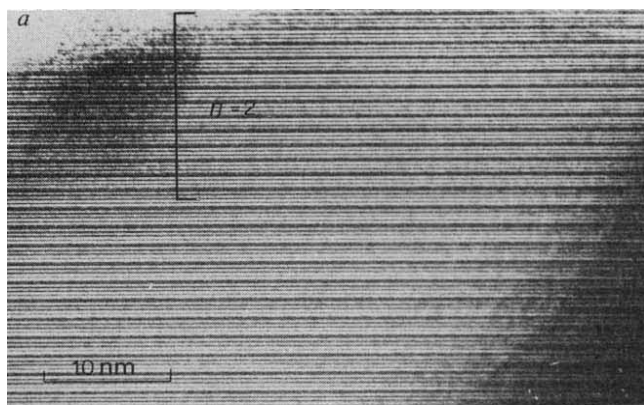


Fig. 3 Electron micrographs *a*, showing $\text{La}_4\text{Cu}_3\text{O}_{10}$ containing a band of $\text{La}_3\text{Cu}_2\text{O}_7$, $n=2$; *b*, $\text{La}_4\text{Cu}_3\text{O}_{10}$ isolated lamellae of $\text{La}_3\text{Cu}_2\text{O}_7$; *c*, a disordered crystal predominantly containing lamellae of $\text{La}_4\text{Cu}_3\text{O}_{10}$, $n=3$, and $\text{La}_5\text{Cu}_4\text{O}_{13}$, $n=4$, and isolated lamellae of $\text{La}_6\text{Cu}_5\text{O}_{16}$, $n=5$, $\text{La}_7\text{Cu}_6\text{O}_{19}$, $n=6$ and $\text{La}_8\text{Cu}_7\text{O}_{22}$. The lamellae of the $n=3$ base structure are 1.42 nm apart.

phases lie in the composition region between La_2CuO_4 and a phase of the perovskite type represented by the idealized structure shown in Fig. 1d. They thus consist of wider and wider perovskite-type slabs, similar to that illustrated in Fig. 1c, d. We return below to the compositions of these phases and the way in which they differ from the idealized diagrams. As an example, Fig. 2 shows a micrograph of a well-ordered fragment of a crystal of the $n=3$ phase $\text{La}_4\text{Cu}_3\text{O}_{10}$. Under the imaging conditions used in the micrograph, the broad white bands represent the regions between the perovskite slabs, which are the sites of the La atoms. The dotted regions between these strips are the

perovskite regions. It is tempting to interpret the dots as showing the centres of the CuO_6 octahedra in the phase, that is, the Cu atoms. Such an interpretation is in accord with the structural information available, but would need to be confirmed by comparison with calculated images before the interpretation could be convincingly supported. This work is in progress.

In addition to La_2CuO_4 , the $n = 1$ phase, and $\text{La}_4\text{Cu}_3\text{O}_{10}$, the $n = 3$ phase shown in Fig. 2, well-ordered examples of the phases $n = 2, 4$ and 5 have been identified. In many cases the crystallites examined appeared to be very well-ordered. However, in common with other non-stoichiometric layer phases, intergrowth of 'wrong' lamellae were frequently seen. If these lamellae are interpreted as being slabs of intergrowth phases rather than simply stacking errors then compounds up to $n = 7$ have been characterized. Typical examples of such disordered micrographs are shown in Fig. 3. In examining these micrographs it is useful to point out that the contrast at the edge of the crystal is very sensitive to the actual imaging conditions used. No significance can be attributed to whether the fringes are white or black, but the heavier of these fringes in all cases corresponds to the layer of structure between the perovskite slabs, that is, the La-containing regions. Thus in Figs 2 and 3c these regions are imaged as broader white lines whereas in Fig. 3a, b, they are imaged as broader black lines. In the thicker parts of the crystal the contrast becomes more complex and should not be interpreted in this way.

The results show that a series of perovskite- K_2NiF_4 structure phases exist in the La-Cu-O system. These can be thought of as intergrowths between La_2CuO_4 and a perovskite oxide of formula LaCuO_3 . In the former of these oxides the Cu is totally in the Cu^{2+} state if the phase is stoichiometric while in the latter it exists entirely as Cu^{3+} . The phases therefore contain both Cu^{2+} and Cu^{3+} , which is the condition forced on La_2CuO_4 by doping with Ba and Sr and which is supposed to be of critical importance in the superconducting phenomenon. It is thus possible each of these new phases will be a high-temperature superconducting phase in its own right. Doping each member of the homologous series with Ba or Sr could also produce superconducting phases if the ratio of Cu^{2+} to Cu^{3+} is not favourable in the prepared materials.

With respect to the Cu^{2+} to Cu^{3+} ratio it should be pointed out that this will depend on the preparation conditions used. The phases are non-stoichiometric and oxygen loss and uptake will be relatively easy; this behaviour is facilitated by the ability of Cu^{2+} to adopt a square planar configuration⁷. Thus the conductivity of any particular sample will be sensitive to the preparation conditions used, particularly the oxygen partial pressure over the reactants. The fact that the series of phases in our preparations are unstable above 900 °C may well be due to this factor. Hence preparation at lower temperatures or under enhanced oxygen partial pressures would be expected to expand the range over which the homologous series exists, while operating at lower oxygen pressures may well decrease the existence range.

The full extent of the series has yet to be determined, but our preliminary results suggest that it is more extensive than that in the Sr-Ni-O and Sr-Ti-O systems. In addition, the exact oxygen stoichiometry of these phases has not yet been confirmed by chemical analysis. Further experiments to investigate this and the effects of doping with other ions will be reported in the future.

Chemical modification of n-GaAs electrodes with Os^{3+} gives a 15% efficient solar cell

Bruce J. Tufts, Ian L. Abrahams, Patrick G. Santangelo, Gail N. Ryba, Louis G. Casagrande & Nathan S. Lewis*

Department of Chemistry, Stanford University, Stanford, California 94305, USA

The minimization of interfacial recombination losses is a key factor in the operation of any semiconductor-based solar-energy-conversion device, including solid-state junctions, semiconductor/liquid junctions and colloidal suspensions of semiconductors. A frequently cited advantage of semiconductor/liquid junctions is the ability to manipulate surface recombination rates by chemical reactions¹⁻⁷. A notable example is the improvement in current-voltage properties of n-GaAs photoanodes which have been exposed to aqueous solutions of Ru^{3+} ions⁸⁻¹⁰. Here we report new surface-modification procedures for GaAs which have produced the most efficient photoelectrochemical cell reported to date. We also report experiments which indicate that the current-voltage improvements in this system are accompanied by increased interfacial hole transfer rates at the GaAs/liquid interface.

When (100) oriented, matte-etched⁹, n-GaAs electrodes (metal-organic chemical vapour deposited, donor density (N_d) = $1 \times 10^{17} \text{ cm}^{-3}$ of Se) are in contact with the 1.0 M KOH-1.0 M Se^{2-} -0.01 M Se_2^{2-} electrolyte they yield stable, but relatively poor, current-voltage (I - V) properties under simulated solar illumination (Fig. 1)¹⁰. Exposure of the etched n-GaAs to an aqueous 0.010 M solution of $\text{OsCl}_3 \cdot 3\text{H}_2\text{O}$ before evaluation in the KOH- Se^{2-} cell yielded far superior I - V properties (Fig. 1), with open-circuit voltages of 780-810 mV, short-circuit photocurrent densities of 24-26 mA cm^{-2} , fill-factors of 0.65-0.75, and cell efficiencies of $15.0 \pm 1.0\%$. These I - V properties remained constant (within $\pm 5\%$) when the cell was maintained at the maximum power point for $>3,000 \text{ C cm}^{-2}$ of charge passed. Note that only 20 C cm^{-2} would have been required to consume the entire photoactive n-GaAs epilayer (10 μm thick) by a 6-electron decomposition process. These I - V properties represent relative improvements by a factor of 1.2-1.3 over the Ru^{3+} ion treatment reported previously, and result in the most efficient photoelectrochemical solar cells reported to date^{6,7,11-13}.

Controversy exists over the dominant mechanism of I - V improvement for the n-GaAs (Ru^{3+}) system, with some investigators postulating an increased hole transfer rate and others postulating a decrease in the hole recombination rate^{1,2,8-10,14,15}. To investigate whether Os^{3+} chemisorption results in an effective electrocatalyst for Se^{2-} oxidation, we have investigated the behaviour of degenerately doped (10% Sn) In_2O_3 (ITO) electrodes. ITO electrodes exhibited reversible electrochemistry in an aqueous $\text{Fe}(\text{CN})_6^{3-/4-}$ solution; however, they exhibited a high overpotential for Se^{2-} oxidation in 1.0 M KOH/1.0 M Se^{2-} /0.01 M Se_2^{2-} solutions (Fig. 2a). Exposure of the ITO to a 0.010 M solution of aqueous Os^{3+} ions yielded marked reductions in the overpotential for Se^{2-} oxidation, indicating that the chemisorbed Os is acting as an effective electrocatalyst for Se^{2-} oxidation. For small overpotentials, the current density was increased by a factor of $\sim 1,000$ by the Os treatment. Similar behaviour was observed for chemisorbed Ru^{3+} ions at ITO electrode surfaces.

To explore further the possibility of electrocatalysis, we have examined the behaviour of (100) orientated p-GaAs electrodes (acceptor density = $4 \times 10^{17} \text{ cm}^{-3}$, Zn doped) in KOH- Se^{2-} . In this electrolyte, we observed rectification properties from

Received 25 March; accepted 9 April 1987.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189-193 (1986).
2. Chu, C. W., Hor, P. H., Meng, R. L., Gao, L. & Huang, Z. *J. Science* **235**, 567-569 (1987).
3. Jayaraman, K. S. *Nature* **326**, 237 (1987).
4. Longo, J. M. & Racah, P. M. *J. Solid St. Chem.* **6**, 526-531 (1973).
5. Tilley, R. J. D. *J. Solid St. Chem.* **21**, 293-301 (1987).
6. Mohan Ram, R. A., Ganapathi, L., Ganguly, P. & Rao, C. N. R. *J. Solid St. Chem.* **63**, 139-147 (1986).
7. Michel, C. & Raveau, B. *Revue Chim. miner.* **21**, 407-425 (1984).

* To whom correspondence should be addressed.

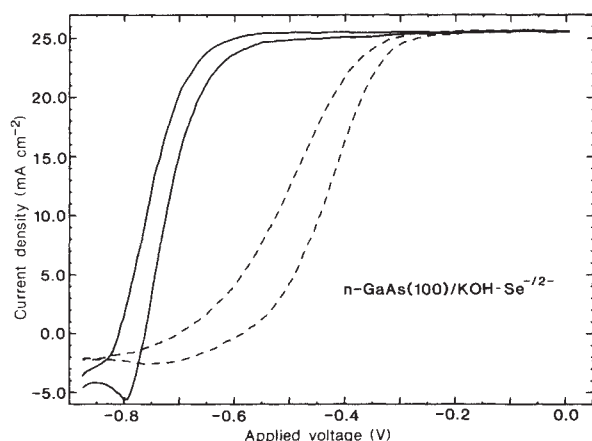


Fig. 1 Photocurrent-voltage properties of n-GaAs anodes in 1.0 M KOH/1.0 M K_2Se /0.01 M K_2Se_2 electrolyte. The electrode was 0.0726 cm² in exposed area, and consisted of a 10- μ m thick n-GaAs epilayer on a conducting n⁺-GaAs substrate. Illumination was provided from a calibrated ELH-type W-halogen bulb¹⁷ to the level of 88 mW cm⁻²; the solutions were magnetically stirred and *I*-*V* data (50 mV s⁻¹ scan rate) were obtained potentiostatically with the n-GaAs as the working electrode. The reference electrode was a Pt wire poised at the equilibrium cell potential and the counterelectrode was a Pt foil. (---), Surface etched to a matte finish with 1:1 $H_2SO_4:H_2O_2$ for 30 s, then rinsed with H_2O . (—), Surface etched to a matte finish, rinsed with H_2O , immersed for 60 s in aqueous ($pH=2$, HCl) 0.010 M $OsCl_3 \cdot 3H_2O$ and rinsed with H_2O . No correction has been made for any optical absorption or reflection losses, and efficiency is calculated as the maximum product of $I \times V$ divided by the total optical power incident on the cell.

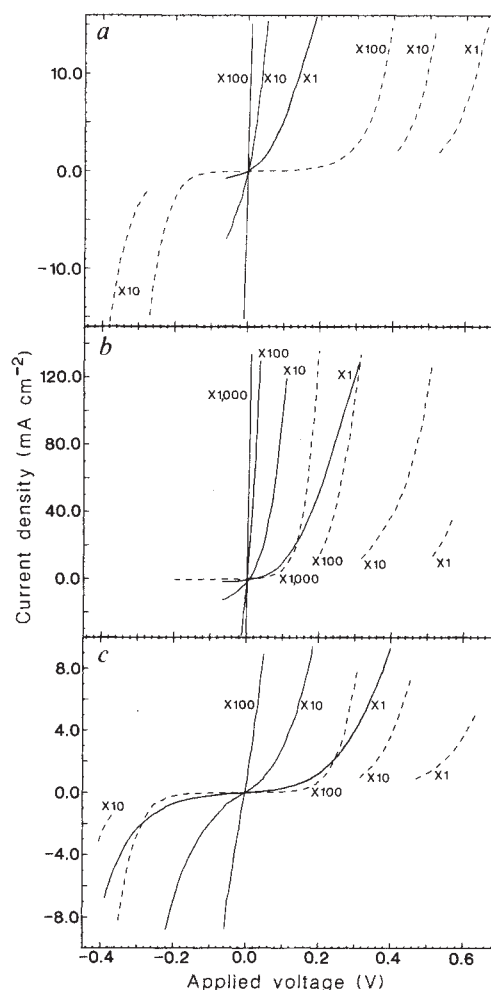


Fig. 2 Current-voltage behaviour (20 mV s⁻¹) of various electrode surfaces in 1.0 M KOH/1.0 M K_2Se /0.01 M K_2Se_2 without (---) and with (—) exposure to 60 s of 0.010 M aqueous ($pH=2$ with HCl) $OsCl_3 \cdot 3H_2O$. The reference electrode was a Pt wire immersed in the solution and poised at the equilibrium cell potential; the counterelectrode was a Pt foil. The units which accompany the individual curves indicate the appropriate increase in sensitivity of the current scale. *a*, *I*-*V* behaviour of 10% Sn doped In_2O_3 electrodes. *b*, *I*-*V* behaviour of etched (4:1:1 $H_2SO_4:H_2O_2:H_2O$ for 10 s, then H_2O rinse) p-GaAs (100) surfaces in the dark without and with exposure to Os^{3+} . *c*, *I*-*V* behaviour of degenerately doped n⁺-GaAs anodes in the dark (same etching procedure as in (*b*) without and with exposure to Os^{3+}).

p-GaAs (Fig. 2*b*). However, when the p-GaAs was exposed to Os^{3+} ions before contact with the KOH- $Se^{-/2-}$ cell, completely different *I*-*V* behaviour was observed, and large increases in the cathodic limiting current (up to the mass transport limit) and dramatic reductions in the overpotential needed to oxidize Se^{2-} were apparent. This behaviour is consistent with an increase in the interfacial hole transfer rate after chemisorption of Os. However, the increased currents could not be produced by a decrease in the surface recombination rate, because the elimination of surface recombination would decrease both the dark anodic and dark cathodic currents at all potentials.

The increased hole-transfer rate for Se^{2-} oxidation could result from a shift in the flat-band potential of the GaAs, or could result from the introduction of electrocatalytic states located near the valence band in energy. Information regarding the electrocatalytic properties of n-GaAs (Os^{3+}) surfaces can be obtained by use of degenerately doped n⁺-GaAs anodes, because a shift in flat-band potential should not be of major consequence to the *I*-*V* properties for degenerately doped material. For n⁺-GaAs samples ($N_d > 5 \times 10^{18}$ cm⁻³, Se-doped) which had been exposed to Os^{3+} ions, we observed a large decrease in the overpotential necessary to oxidize Se^{2-} (Fig. 2*c*), and therefore we conclude that the metal ion treatment acts as an efficient electrocatalyst for Se^{2-} oxidation at the n-GaAs/KOH- $Se^{-/2-}$ junction. Although it is possible that electron injection is the dominant exchange current process for n⁺-GaAs(Os^{3+}) electrodes whereas hole emission is the dominant photocurrent exchange process for the n-GaAs(Os^{3+}) oxidations, the observation of improved anode properties at all GaAs surfaces is certainly consistent with the electrocatalysis of Se^{2-} oxidation observed at Os^{3+} -treated ITO surfaces.

Finally, we have investigated the effect of Os^{3+} treatment on the high-level surface recombination velocity of the (100) n-GaAs/air interface. A reduction in the surface-state density or in the carrier-trapping cross-sections could produce an increase in the intensity of the bulk radiative luminescence. We observed no change in the steady-state luminescence intensity

(632.8 nm excitation from a focused 5 mW He/Ne laser; luminescence at 1.42 eV) of an n-GaAs sample ($N_d = 2 \times 10^{17}$ cm⁻³, Se-doped, hole diffusion length $> 5 \mu$ m) which had been subjected to the following surface treatments: etched (4:1:1 $H_2SO_4:H_2O_2:H_2O$); etched and exposed to Os^{3+} ; etched and exposed to Os^{3+} and electrochemically cycled in KOH- $Se^{-/2-}$; etched and immersed in KOH- $Se^{-/2-}$ and then exposed to Os^{3+} . This behaviour contrasts with the reported decrease in the n-GaAs/air-surface recombination velocity upon chemisorption of Ru^{3+} ions¹⁶.

We have demonstrated that chemisorption of metal ions can yield highly beneficial improvements in the *I*-*V* behaviour of semiconductor/liquid junctions. We have also demonstrated that electrocatalytic oxidation of Se^{2-} can be an important feature of the *I*-*V* properties of photoanodes in the KOH- $Se^{-/2-}$ electrolyte, and that surface-modification procedures which increase the hole-transfer rate in this electrolyte can effect improved solar cell performance. It is possible that Os^{3+} chemisorption could passivate recombination at grain bound-

aries even though we observe no decrease in surface-recombination velocity at crystalline n-GaAs/air junctions, and such effects are being investigated at present.

We acknowledge the NSF for support of this work. N.S.L. also acknowledges support as a Dreyfus-Teacher Scholar and as an Alfred P. Sloan Foundation Fellow. We also thank Dr C. R. Lewis of Varian Associates, Palo Alto, California, for a supply of GaAs samples.

Received 29 December 1986; accepted 26 February 1987.

- Heller, A. *Am. chem. Soc. Symp. Ser.* **146**, 57-77 (1981).
- Heller, A. *Accs chem. Res.* **14**, 154-162 (1981).
- Bard, A. J. *J. phys. Chem.* **86**, 172-177 (1982).
- Kiwi, J., Kalyanasundaram, K. & Gratzel, M. in *Structure and Bonding* Vol. 49 (ed. Jorgensen, K.) 37-125 (Springer, Berlin, 1982).
- Wrighton, M. S. *Accs chem. Res.* **12**, 303-310 (1979).
- Rajeshwar, K. *J. appl. Electrochem.* **15**, 1-22 (1985).
- Peter, L. M. *Electrochemistry* **9**, 66-100 (1978).
- Parkinson, B. A., Heller, A. & Miller, B. *Appl. Phys. Lett.* **33**, 521-523 (1978).
- Parkinson, B. A., Heller, A. & Miller, B. *J. electrochem. Soc.* **126**, 954-960 (1979).
- Chang, K. C., Heller, A., Schwartz, B., Menezes, S. & Miller, B. *Science* **196**, 1097-1099 (1977).
- Parkinson, B. A. *Accs chem. Res.* **17**, 431-437 (1984).
- Lewis, N. S. *A. Rev. Mater. Sci.* **14**, 95-117 (1984).
- Tenne, R. & Wold, A. *Appl. Phys. Lett.* **47**, 707-709 (1985).
- Allongue, P. & Cachet, H. *J. electrochem. Soc.* **131**, 2861-2868 (1984).
- Ludwig, M., Heymann, G. & Janietz, P. *J. Vac. Sci. Technol.* **B4**, 485-492 (1986).
- Nelson, R. J. *et al. Appl. Phys. Lett.* **36**, 76-78 (1980).
- Gronet, C. M. & Lewis, N. S. *J. phys. Chem.* **88**, 1310-1317 (1984).

A light-variation insensitive high efficiency solar cell

Stuart Licht*, Gary Hodes, Reshef Tenne & Joost Manassen

Department of Materials Research, Weizmann Institute of Science, Rehovot 76100, Israel

We report a photoelectrochemical solar cell (PEC) with *in situ* storage, operating at 11.3% overall solar to electrical conversion efficiency, with a generated power insensitive to daily fluctuations in solar radiation. A primary advantage of PECs, compared to photovoltaic solar cells, had been their proposed capacity for *in situ* storage by the addition of a single electrochemical half cell¹. Several systems combining photoelectrochemical conversion with electrochemical storage have been presented with maximum efficiencies of approximately 3 per cent. A series of polysulphide solution effects²⁻⁵ resulted in a high efficiency caesium polysulphide n-Cd(Se,Te) single crystal PEC⁶. As will be shown, this leads to an autonomous cell with sufficient voltage to bypass the previous necessity for bipolar dual photoelectrode storage, allowing single photoelectrode induced electrochemical storage, with a high continuous output.

In principle, Cd(Se, Te) with a band gap of ~1.4 eV (ref. 7), when immersed in a polysulphide electrolyte of redox potential -0.5 V versus the standard hydrogen electrode (SHE), should be sufficiently energetic to drive stannous sulphide reduction with a redox potential of -1.0 V versus SHE. In practice, only a fraction of the band gap had been accessed in these PECs. The photovoltage is further diminished with increased photocurrents as reflected by fill factors generally <0.7. Hence at maximum power, potentials of typically ~-0.9 V versus SHE had been generated. This is insufficient to charge a tin sulphide storage system. However, a modified polysulphide electrolyte PEC was recently shown to deliver -1.1 V versus SHE at a maximum power conversion efficiency of 12.7%⁶. In this paper, the enhanced potential is proved to be sufficient to drive a storage component directly.

* To whom correspondence should be addressed: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

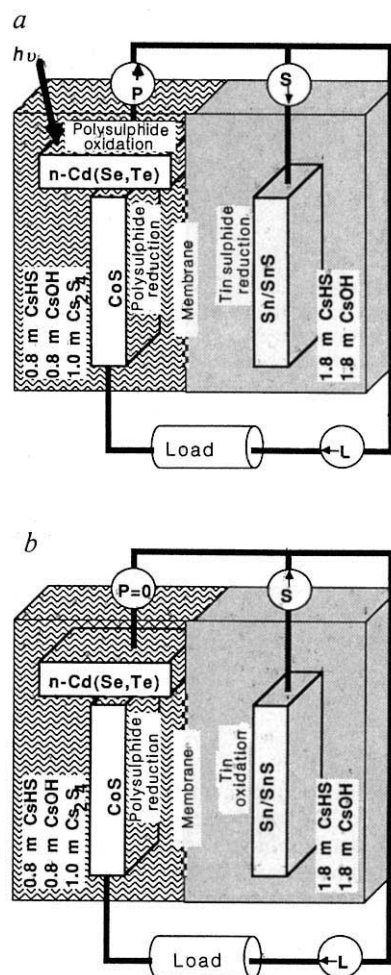
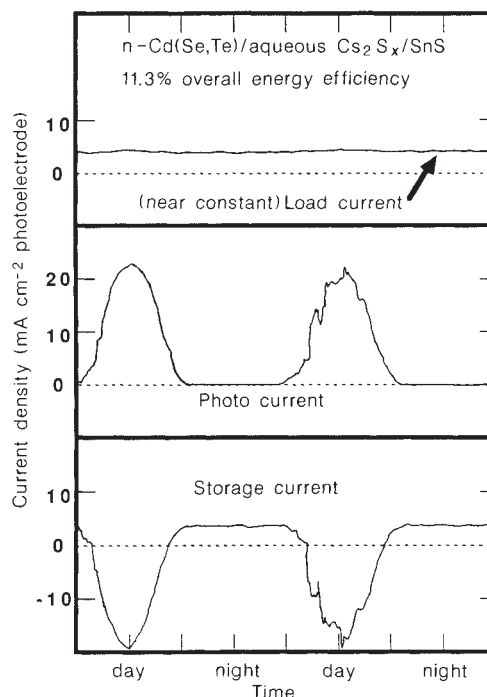


Fig. 1 The n-Cd(Se, Te)/aqueous $\text{Cs}_2\text{S}_x/\text{SnS}$ solar cell. P, S and L indicate the direction of electron flow respectively through the photoelectrode, tin electrode and external load, a, in an illuminated cell and b, in the dark. For electrolytes, m represents molal.

The cell we have studied is presented schematically in Fig. 1. Illumination of the n-type semiconductor with radiation of greater than band gap energy, induces charge separation within the semiconductor space charge layer. As indicated in Fig. 1a, this creates holes for polysulphide oxidation at the semiconductor surface and drives electron transfer both through an external load and also into electrochemical storage by reduction of SnS to metallic tin. In the dark, Fig. 1b, the potential drop below that of tin sulphide reduction induces spontaneous oxidation of tin, and electron flow through the external load. The inherent resistance of the photoelectrode to polysulphide reduction, minimizes dark losses through the photoelectrode. The nominal relative concentration of sulphur and sulphide within the photoelectrode and tin electrode compartments varies with degree of cell charge. Note that independent of illuminating conditions, electrons are driven through the load in the same direction, ensuring continuous power.

A 0.080-cm^2 n-CdSe_{0.65}Te_{0.35} single crystal was mounted with the 0001 plane exposed, and pre-treated with mechanical polishing and chemical etching, followed by rinsing and photoelectrochemical etching⁸, as previously described⁶. A 3-cm^2 thin film CoS counter electrode, electrocatalytic to the polysulphide redox reaction⁹, was prepared by electrodeposition of cobalt onto brass with subsequent electrochemical oxidation to CoS in polysulphide solution. A 3-cm^2 Permion 1010 membrane, a sulphonated copolymer of styrene and teflon, allowed electrolytic transport

Fig. 2 Two days outdoor operation of an n-Cd(Se, Te)/aqueous $\text{Cs}_2\text{S}_x/\text{SnS}$ solar cell driving a 1,500 Ω load. The cell contains an 0.080-cm² single-crystal n-Cd(Se, Te) photoelectrode and a 3-cm² thin film CoS counter electrode immersed in a 0.8 molal CsHS, 0.8 molal CsOH and 1.0 molal Cs_2S_4 electrolyte separated by Permion 1010 sulphonated teflon/styrene membrane from a Sn electrode immersed in an 1.8 molal CsHS and 1.8 molal CsOH electrolyte. Load, photo and storage currents are measured as indicated in Fig. 1.



between adjacent compartments of a lucite cell. One compartment contained the photoelectrode, counter electrode and 2.5 cm³ of the polysulphide electrolyte. It was separated by the membrane from a compartment containing a mechanically polished 3-cm² tin electrode immersed in 2.5 cm³ of the sulphide electrolyte. Commercial analytical grade aqueous 50% CsOH was reacted quantitatively with H₂S to form CsHS solutions, as previously described³. Dilution and addition of CsOH and sulphur forms the electrolytes with initial chemical concentrations indicated in Fig. 1. In solution a complex equilibrium of the various polysulphide species develops¹⁰.

The cell was first cycled in the dark. This enhances secondary storage by modifying the tin surface morphology, and hence decreasing Sn/SnS oxidation or reduction overpotentials to less than 20 mV cm² mA⁻¹. Unlike previous cells studied, the open-circuit voltage of the charged storage cell fell from its initial value of 0.510 V to a stable value of 0.480 V, reflecting the decreased selectivity of the membrane used in this study.

Under direct illumination, the 0.080-cm² photoelectrode in this cell generated more than 1.5 mA through the 3 cm² SnS electrode, driving tin sulphide reduction while supporting 0.33 mA through the 1,500 Ω load simultaneously at a photogenerated 0.495 V. Spontaneous tin oxidation in the dark at 0.470 V continues to drive the load, with a storage efficiency of:

$$(0.470 \text{ V}/0.495 \text{ V}) \times 100 = 95\%.$$

Under 96.5 mW cm⁻² solar radiation, as measured on an Eppley pyranometer⁶, the photoelectrode generated 23 mA cm⁻² at 0.495 V, yielding a solar to electrical direct energy conversion efficiency of:

$$(23 \text{ mA cm}^{-2} \times 0.495 \text{ V}/96.5 \text{ mW cm}^{-2}) \times 100 = 11.8\%.$$

This efficiency is lower than the previously⁶ reported efficiency (without storage) of 12.7% due to increased light absorbance of the lucite compared to a glass-walled cell and the tin reduction potential, which causes the cell to operate at ~100 mV below the maximum power point voltage of the cell.

Stability tests in PECs (without storage) for single-crystal and thin-film n-Cd(Se, Te) cells both using the same photoelectrolyte employed in this study, have been reported^{6,11}. Two days of

outdoor operation for the cell, summarized in Fig. 1 and described here, are presented in Fig. 2. As is evident in Fig. 2, there is a continuous and nearly constant power generated through the external load, independent of fluctuations in incident radiation. Approximately one fifth of solar radiation is directly used through the load at 11.8% efficiency. The balance is regenerated by electrochemical storage at 11.2% efficiency (representing 95% retrieval of the original efficiency). Combined over the load, these yield an overall solar-to-electrical load conversion efficiency of 11.3%.

The reported system demonstrates, in the case of an n-type cadmium chalcogenide/aqueous polysulphide PEC with *in situ* tin sulphide storage, that it is the photoconversion component, not the electrochemical storage component, which limits overall efficiency. Thin film n-Cd(Se, Te)/aqueous polysulphide PECs (without storage) have demonstrated efficiencies of greater than 7 per cent^{12,13} or stabilities of one year¹¹. A thin film photoelectrode PEC device would be economically more viable than a comparable single crystal device. It is hoped that this study will add impetus to the search for viable high efficiency thin-film PEC systems.

We thank R. Triboulet, J. Rioux and C. Levy-Clement of the Laboratoire de Physique des Solids, Meudon, France, for the n-Cd(Se, Te) crystal employed in the study, and Menashe Lev of the Weizmann Institute of Science for his technical services. This work was supported in part by the Solar Energy Research Institute under subcontract XE-2-02096-01.

Received 5 December 1986; accepted 4 February 1987.

1. Manassen, J., Hodes, G. & Cahen, D. *J. electrochem. Soc.* **124**, 532 (1977).
2. Licht, S. & Manassen, J. *J. electrochem. Soc.* **132**, 1076-1081 (1985).
3. Licht, S., Tenne, R., Flaisher, H. & Manassen, J. *J. electrochem. Soc.* **133**, 53-59 (1986).
4. Licht, S., Manassen, J. & Hodes, G. *J. electrochem. Soc.* **133**, 272-277 (1986).
5. Licht, S. & Manassen, J. *J. electrochem. Soc.* **133**, 277-280 (1986).
6. Licht, S. *et al. Appl. Phys. Lett.* **46**, 608-610 (1985).
7. Tai, H., Nakashima, S. & Hori, S. *Status Solidi* **30**, K115-119 (1975).
8. Tenne, R. & Hodes, G. *Surf. Sci.* **135**, 453-478 (1983).
9. Hodes, G., Manassen, J. & Cahen, D. *J. electrochem. Soc.* **127**, 544-549 (1980).
10. Licht, S., Hodes, G. & Manassen, J. *Inorg. Chem.* **25**, 2486-2489 (1986).
11. Licht, S. *J. phys. Chem.* **90**, 1096-1099 (1986).
12. Hodes, G. *Nature* **285**, 29-30 (1980).
13. Russak, M. A. & Creter, C. *J. electrochem. Soc.* **131**, 556 (1984).

First direct radiometric dating of Archaean stromatolitic limestone

S. Moorbath*, P. N. Taylor*, J. L. Orpen†, P. Treloar‡§ & J. F. Wilson†

* Department of Earth Sciences, Oxford University, Parks Road, Oxford OX1 3PR, UK

† Geology Department, University of Zimbabwe, PO Box MP 167, Mount Pleasant, Harare, Zimbabwe

‡ Department of Geology and Physical Sciences, Oxford Polytechnic, Headington, Oxford OX3 0BP, UK

Here we report the first direct dating of the depositional age of a sedimentary carbonate rock using long-lived radioactive decay schemes. The Mushandike stromatolitic limestone, from the Masvingo (formerly Fort Victoria) greenstone belt of southern Zimbabwe^{1,2}, yields a Pb/Pb isochron age of $2,839 \pm 33$ Myr. Published claims³ for an age of ~ 3.5 Gyr for the Mushandike limestone were based on its supposed intrusion by the Mushandike granite, for which a Rb–Sr whole-rock age of $3,445 \pm 260$ Myr had been reported⁴. New Rb–Sr and Pb/Pb isotope data for the Mushandike granite indicate a probable emplacement age of $\sim 2,900$ Myr, and new field evidence (J.L.O., P.T. & J.F.W., in preparation) suggests that the granite forms part of the basement on which the limestone was deposited. Thus the new dates do not support an age of ~ 3.5 Gyr for the algal stromatolites at Mushandike. The Pb/Pb isochron method may have wide applicability to dating the time of deposition of Precambrian fossiliferous and unfossiliferous limestones, and thus to the problem of the antiquity of life.

Three different ages of greenstone belt development have been recognized in the Zimbabwe craton, namely the Sebakwian Group at ~ 3.5 Gyr, the lower part of the Bulawayan Group (the Lower Greenstones) at ~ 2.9 Gyr, and the upper part of the Bulawayan (the Upper Greenstones), with the associated Shamvaian Group, at ~ 2.7 Gyr (ref. 2). Whereas most greenstone belt units can be assigned to one of these major stratigraphic subdivisions, some are difficult to allocate. These include the sedimentary remnant which contains the Mushandike stromatolitic limestone (MSL), originally regarded as Bulawayan but later grouped with the Sebakwian³.

The MSL outcrops within the Mushandike National Park, 25 km west of Masvingo (Fig. 1). The park consists of a northern 'granitic' area and a southern 'greenstone' area. In the north, migmatitic gneisses are intruded by the Mushandike granite (MG), for which Hickman⁴ obtained a Rb–Sr whole-rock isochron date of $3,445 \pm 260$ Myr. In the south, on Gwenya Hill, banded iron formation (BIF) and ferruginous siltstones of the Gwenya formation, containing small, highly deformed lenses of limestone, unconformably overlie the MG. The Gwenya formation can be traced among much of the northern margin of the Masvingo greenstone belt and is overlain to the south by a thick sequence of basalts in a succession typical of the lower part of the Upper Greenstones.

The sedimentary remnant which includes the MSL and associated pelites and amphibolites is deformed in a complex fashion. Nowhere is it in direct contact with the Gwenya formation and it is separated from it by talc schists and serpentized peridotites of the Mashaba Igneous Complex². Orpen and Wilson³ assigned the remnant to the Sebakwian Group based on the interpretation of a late low-pressure, andalusite-forming metamorphism of the rocks as a contact effect of the MG intrusion. More recent field interpretation (P.T., in preparation), based on the observation that the andalusite porphyroblasts overprint semi-ductile micaceous shear zones which cut the MG, indicates that the MG was not the heat source for the metamorphism.

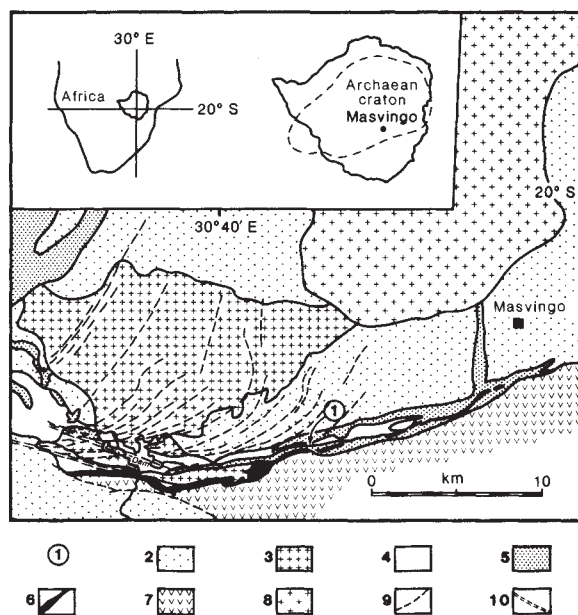


Fig. 1 Geological map of the Mushandike area. 1, Stromatolite locality; 2, basement gneisses; 3, Mushandike granite; 4, Mushandike stromatolitic limestone and associated sediments and amphibolites; 5, Mashaba igneous complex; 6, Gwenya Formation, basal Upper Greenstones; 7, Upper Greenstone volcanics; 8, Chilimanzi granite; 9, curved shear zones; 10, ESE-trending ductile shears.

Sr and Pb isotope analyses were carried out on a fully automated VG–Isomass 54E mass spectrometer, following published procedures⁵. Sr and most Pb samples were extracted using dissolution and ion-exchange techniques, but some Pb samples were prepared by a double-electrodeposition method⁶. Pb concentrations of several samples were determined by mass-spectrometric isotope dilution. Rb/Sr ratios were determined by standard X-ray fluorescence methods on pressed powder pellets⁷. Uncertainties (1σ) in Sr and Pb isotope ratios are $\sim 0.01\%$ and $\sim 0.1\%$ respectively, and $\sim 1\%$ in Rb/Sr ratios. Isochron regressions were carried out using the method of York⁸, and all ages were calculated or recalculated using the decay constants listed by Steiger and Jäger⁹. All age and isotope results are quoted at the 95% (2σ) confidence level.

MG samples were collected mostly from the northern and central parts of the intrusion. An 18-point whole-rock Rb–Sr isochron (mean squares of weighted deviates, MSWD = 30) on

Table 1 Analytical data for Mushandike granite

Sample no.	$^{87}\text{Rb}/^{86}\text{Sr}$	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
28A	1.625	0.78427	23.062	17.320	41.151
28B	0.926	0.75051	23.842	17.378	40.905
29	3.104	0.83241	64.541	27.119	77.450
30A	0.536	0.73176	22.623	17.225	37.789
30B	0.423	0.72802	21.617	16.980	37.852
30C	0.596	0.73566	24.213	17.576	40.729
31	0.775	0.74123	16.990	15.589	34.442
36	0.673	0.74063	18.532	16.175	39.246
37A	0.920	0.74924	18.581	16.087	39.430
37B	1.920	0.79297	22.089	17.169	41.793
38	1.018	0.75102	24.422	17.665	40.561
39	0.734	0.74196	22.762	17.325	40.963
41	1.488	0.77207	26.193	18.071	40.668
42	2.550	0.82497	17.929	15.847	37.779
43	0.894	0.74490	23.951	17.536	39.978
44	0.951	0.74918	22.301	17.186	39.953
46	1.187	0.75523	20.898	16.946	36.563
47	1.766	0.78064	22.114	17.199	39.799

§ Present address: Geology Department, Imperial College, Prince Consort Road, London SW7 2AZ, UK.

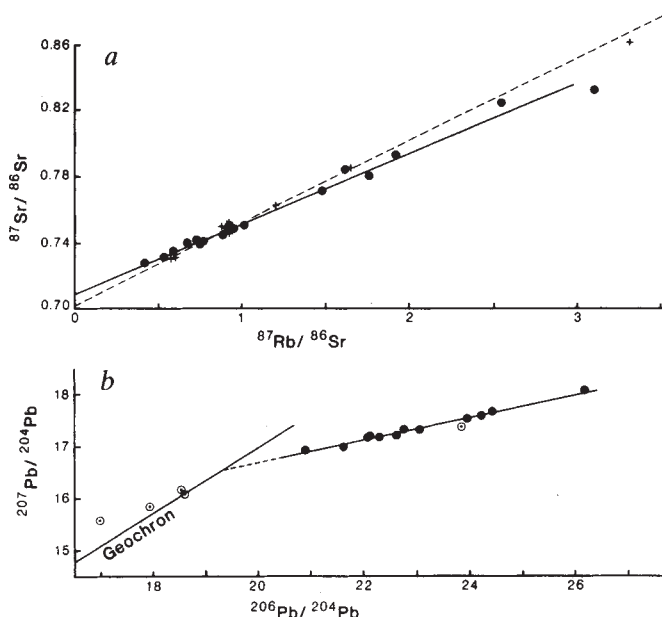


Fig. 2 Isotope data plots for Mushandike granite. *a*, Whole-rock Rb-Sr errorchron (filled circles): age = $2,917 \pm 171$ Myr; $Sr_i = 0.7098 \pm 0.0022$; MSWD = 30. Dashed line and crosses represent errorchron and data points of Hickman⁴: age = $3,445 \pm 260$ Myr; $Sr_i = 0.7018 \pm 0.0031$; MSWD = 15. *b*, Whole-rock Pb/Pb errorchron: age = $2,946 \pm 130$ Myr; model $\mu_1 = 9.7$; MSWD = 5.2. Only filled circle data points have been used in the regression calculation.

the MG (Table 1, Fig. 2a) yields a date of $2,917 \pm 171$ Myr, with an initial $^{87}\text{Sr}/^{86}\text{Sr}$ (Sr_i) value of 0.7098 ± 0.0022 . This contrasts strongly with the 7-point Rb-Sr errorchron (MSWD = 15) date of $3,445 \pm 260$ Myr, with $Sr_i = 0.7018 \pm 0.0031$, of Hickman⁴ (regression line shown in Fig. 2a; MSWD and Sr_i computed from Hickman's data). A 12-point whole-rock Pb/Pb errorchron (MSWD = 5.2) on the MG (Table 1, Fig. 2b, filled circles) gives a date of $2,946 \pm 130$ Myr with a model μ_1 (source $^{238}\text{U}/^{204}\text{Pb}$) of 9.7. Six analytical points have not been used in the regression calculation (Fig. 2b, open circles; one very radiogenic point omitted) solely because of the lack of fit to the regression line.

The new Rb-Sr and Pb/Pb errorchron dates agree within

error and are provisionally interpreted as the intrusive age. Scatter about both regressions in excess of analytical error, as well as the aberrant points in Fig. 2, may be attributable to initial isotope heterogeneity of magma resulting from incomplete mixing of anatectic melt segregations. Thus the high Sr_i and μ_1 values relative to a contemporaneous mantle source (~ 7.8 – 8.5 for late Archaean calc-alkaline granitoids from Zimbabwe, S.M. and P.N.T., unpublished data) suggest that the MG magma was produced by partial melting of an ancient crustal protolith. This is supported by a Sm-Nd model age (assuming no fractionation of Sm-Nd during anatexis) of ~ 3.5 Gyr for a single sample of MG¹⁰, suggesting that the protolith of the MG was similar to the dated ~ 3.5 Gyr-old regional basement gneisses¹¹.

Cores of MSL were drilled from the outcrop previously described by Orpen and Wilson³, ~ 9 km east of Mushandike Dam Wall in the northern margin of the Masvingo greenstone belt. Here we report a Pb isotope study on a single core (no. 2) of MSL 25 cm in length and 2.5 cm in diameter (Table 2). The core was split into 23 approximately equal-sized fragments, and ~ 0.5 g pieces were taken for Pb extraction as follows. Seven samples were subjected to total rock dissolution first in HCl, then in HNO_3 and HF. Nine samples were leached with warm, 3 M HCl to obtain acid-soluble fractions. The insoluble (silicate) residues from five of these were then digested in HNO_3 and HF. The aim of these procedures was to test for isotope contrast and disequilibrium between Pb in the carbonate minerals and any ancient crustal Pb in detrital components within the MSL.

Figure 3 shows a plot of $^{207}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ for all 21 Pb samples, defining an errorchron (MSWD = 75) date of $2,839 \pm 33$ Myr. No significantly different subsets of isotope data can be extracted from the overall array. This wide range of Pb isotope ratios (and corresponding U/Pb ratios) far exceeds that found for most igneous and metamorphic rock units. Evidently, centimetre-sized domains of MSL have remained virtually closed systems to migration of U and Pb for the past ~ 2.84 Gyr. We interpret the errorchron date (Fig. 3) as close to the age of deposition of the MSL, although resetting due to the low pressure, andalusite-forming metamorphic event cannot be ruled out. Abell *et al.*¹² conclude from O- and C-isotope measurements that the MSL had a metamorphic history involving volatile loss at 200°C or less, and may also have undergone exchange with groundwaters to a depth of several centimetres. In contrast, on the basis of amphibole-plagioclase compositions (Treloar, unpublished data) in adjacent metabasalts, as well as the presence of the andalusite-cordierite pair in nearby metapelites and shear zones, we estimate the temperature of low-pressure regional metamorphism to be ~ 450 – 500°C . The stromatolitic limestones occurs in a block in which the carbonate is largely unrecrystallized. The block is a low-strain augen, bounded by shear zones with intense carbonate recrystallization. These features suggest that fluid exchange between MSL and surrounding rocks during metamorphism may have been minimal, with fluid flow being channelled along shear zones.

To develop the large range of present-day Pb isotope compositions in the MSL (Fig. 3) would require *in situ* μ values from ~ 20 to ~ 113 since 2,839 Myr, assuming a plausible initial Pb isotope composition. Measured Pb concentrations in six of the MSL samples range from 2.5 to $16.7 \mu\text{g g}^{-1}$, corresponding to initial Pb concentrations of ~ 1.7 to $\sim 12.3 \mu\text{g g}^{-1}$. Between $\sim 25\%$ and $\sim 60\%$ of the Pb in MSL samples today is radiogenic, developed since 2,839 Myr. These data imply present-day U concentrations in the MSL between ~ 1.0 and $\sim 10.4 \mu\text{g g}^{-1}$, or initial concentrations of ~ 1.7 to $\sim 17.3 \mu\text{g g}^{-1}$.

The single-stage model μ_1 value of 12.4 for the MSL (Fig. 3) is much higher than the range of ~ 7.8 to ~ 8.5 determined from whole rock Pb/Pb isochrons of late Archaean igneous rocks with presumed mantle-derived protoliths from Zimbabwe (S.M. and P.N.T., unpublished data). The high μ_1 value for MSL is most plausibly interpreted as the result of a short-lived high- μ stage of Pb isotope evolution immediately before the closure of

Table 2 Analytical data for Mushandike stromatolitic limestone

Sample no.	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
02T	104.893	35.415	84.673
04T	45.253	23.305	51.654
06T	48.296	23.883	53.737
10T	44.577	22.784	49.756
12T	68.829	27.478	61.235
14T	54.852	24.817	58.231
20T	66.432	26.950	64.312
02S	140.499	41.695	116.210
04S	131.160	40.201	103.374
08S	45.987	23.253	54.170
10S	31.468	20.236	45.445
12S	61.833	26.189	62.776
14S	99.306	33.381	88.349
16S	51.808	24.401	57.402
20S	115.047	37.573	88.544
22S	39.268	21.600	50.011
04R	91.646	32.570	63.827
06R	81.030	29.728	59.113
10R	88.324	31.814	53.150
16R	130.087	40.323	72.140
20R	136.287	41.721	81.101

T, total rock dissolution; S, acid-soluble fraction; R, residue after acid leaching.

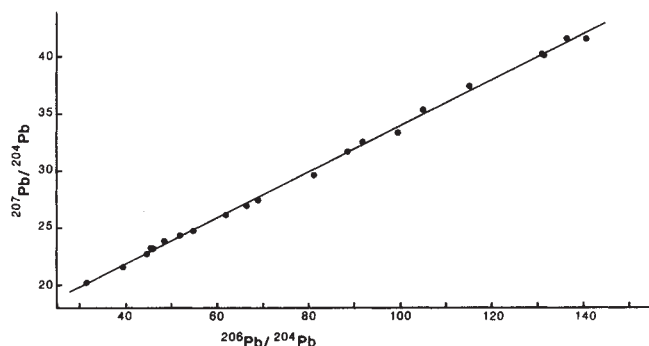


Fig. 3 Pb/Pb errorchron for Mushandike stromatolitic limestone, drillcore no. 2: age = $2,839 \pm 33$ Myr; model $\mu_1 = 12.4$; MSWD = 75.

the MSL U-Pb systems at $\sim 2,839$ Myr. Sea water today has a μ value of $\sim 95,000^{13}$, but the very short residence time of ~ 50 yr for Pb in sea water¹³ is insufficient for significant development of radiogenic Pb in the aqueous medium. Consequently a very high- μ stage in sea water cannot account for the large increase in model μ_1 value. Furthermore, the very low concentration of ~ 2 pg g⁻¹ Pb in modern sea water¹³ could not have produced the initial Pb concentrations in the MSL (see above) without an effective process of enrichment of Pb relative to Ca during carbonate deposition. In contrast, with the U concentration of ~ 3.1 ng g⁻¹ characteristic of modern sea water¹³, the initial U content of the MSL (see above) could be a primary depositional feature not requiring enrichment of U relative to Ca during carbonate deposition.

Modern stromatolitic limestones are developed in shallow-marine, intertidal environments. Assuming a similar environment for deposition of the MSL, post-depositional emergence might induce a flux of ground water derived from continental basement through the newly deposited carbonates. Pb concentration in fresh water is likely to be orders of magnitude greater than sea water. (Modern river water has ~ 1 ng g⁻¹ Pb¹³.) Thus, diagenetic recrystallization of carbonates involving continent-derived ground waters as pore fluid might account for the initial Pb concentration of the MSL. With high primary (depositional) U concentrations, the overall time span of diagenesis of the MSL may represent a high- μ stage of sufficient duration to develop enough radiogenic Pb to attain the high model μ_1 value of 12.4 by the time of closure of MSL U-Pb domains at $\sim 2,839$ Myr.

Assuming that the MSL was deposited on or near a basement of MG and that Pb in ground water was derived mainly from MG, the initial Pb isotope composition of MSL could have developed over a time interval of <100 Myr in an environment with $\mu > 170$. This is significantly greater than the range of μ values of ~ 30 to ~ 113 for the MSL, so that a reduction in U/Pb may have occurred between commencement of deposition and cessation of diagenesis (or low-grade metamorphism).

Preliminary Pb isotope data (not tabulated here) on other drill cores have given similar results. Five separate acid-soluble fractions from another core (no. 4) have yielded an isochron (MSWD = 2.1) date of $2,875 \pm 73$ Myr, with a model μ_1 value of 12.5. Single samples from seven different drill cores (excluding the two described above) yield a combined regression date of ~ 2.85 Gyr, but show evidence for significant scatter in model μ_1 values between different cores.

The isotope work shows that the emplacement of the MG and deposition of the MSL occurred between $\sim 2,950$ and $\sim 2,800$ Myr ago, ~ 600 – 700 Myr later in the Archaean than originally claimed^{3,4}. The MSL and its associated sediments are clearly not part of the Sebakwian Group. We regard the Pb/Pb technique as extremely promising for the direct dating of time of deposition of ancient stromatolitic limestones as well as other sedimentary horizons intercalated with, or bracketed by, lime-

stone beds. If the great ranges in U/Pb and corresponding Pb isotope ratios between adjacent small domains turn out to be a common feature of sedimentary carbonates, it may be possible to extend the applicability of the Pb/Pb dating method to limestones through most of the Proterozoic.

We thank Roy Goodwin and Douglas Foster for technical assistance, and Bor-ming Jahn for critical review.

Received 18 December 1986; accepted 27 February 1987.

1. Wilson, J. F. *Rhod. geol. Surv. Bull.* **62** (1968).
2. Wilson, J. F. *Spec. Publ. geol. Soc. S. Afr.* **5**, 1–23 (1979).
3. Orpen, J. L. & Wilson, J. F. *Nature* **291**, 218–220 (1981).
4. Hickman, M. H. *Nature* **251**, 295–296 (1974).
5. Taylor, P. N., Jones, N. W. & Moorbath, S. *Phil. Trans. R. Soc. A* **310**, 605–625 (1984).
6. Arden, J. & Gale, N. H. *Analyt. Chem.* **46**, 2–9 (1974).
7. Pankhurst, R. J. & O'Nions, R. K. *Chem. Geol.* **1**, 127–136 (1973).
8. York, D. *Earth planet. Sci. Lett.* **5**, 320–324 (1969).
9. Steiger, R. H. & Jäger, E. *Earth planet. Sci. Lett.* **26**, 359–362 (1977).
10. Moorbath, S., Taylor, P. N. & Jones, N. W. *Chem. Geol.* **57**, 63–86 (1986).
11. Moorbath, S., Wilson, J. F., Goodwin, R. & Humm, M. *Precamb. Res.* **5**, 229–239 (1977).
12. Abell, P. I., McClory, J., Martin, A., Nisbet, E. G. & Kyser, T. K. *Precamb. Res.* **27**, 385–398 (1985).
13. Taylor, S. R. & McLennan, S. M. *The Continental Crust; its Composition and Evolution* (Blackwell, Oxford, 1985).

Scattering and absorption of visible light by sea ice

R. G. Buckley* & H. J. Trodahl†

* Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Private Bag, Lower Hutt, New Zealand

† Physics Department, Victoria University of Wellington, Private Bag, Wellington, New Zealand

The scattering and absorption of sunlight falling on the sea-ice cover, which every winter doubles the effective area of the Antarctic continent, exerts a strong influence over the weather of the Southern Hemisphere and over the growth of microbial communities in the water column and within the ice itself¹. It is thus important that we develop an understanding of the interaction of light with sea ice, and here we report *in situ* measurements of the diffusive transport of light in the sea ice of McMurdo Sound, Antarctica. Our novel experimental method permits the resolution of the depth dependence of the scattering and allows identification of an isotropic top layer, an anisotropic bulk layer and a strongly absorbing algal layer. The anisotropic scattering exerts a strong influence on the radiation field in and under the ice.

The few *in situ* measurements of the optical properties of sea ice that have been reported have relied on natural light as a

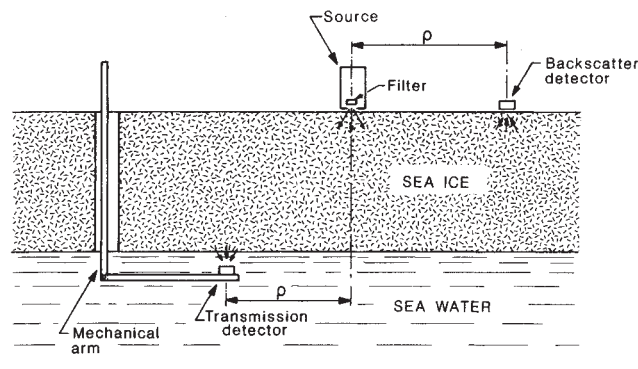


Fig. 1 Schematic diagram of the source and detector layout for both the transmission and backscattering measurement. The flux is measured as a function of the horizontal component of the source-detector separation, ρ .

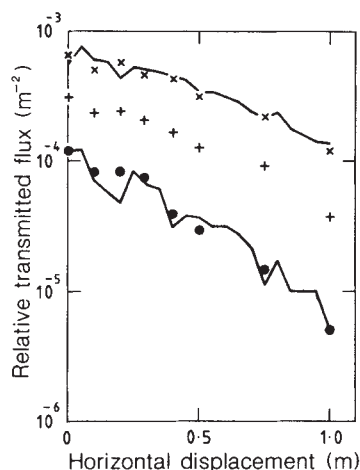


Fig. 2 Plot of the relative transmitted flux against source-detector displacement at site C for $\times$, 500 nm; $+$, 650 nm; $\bullet$, 700 nm. This plot illustrates the fall off in flux with increasing source-detector displacement and the narrowing of the distribution with increasing wavelength. Monte Carlo simulation fits to the 500 nm and 700 nm data are shown as solid lines. The scattering model used in the simulation is described in the text.

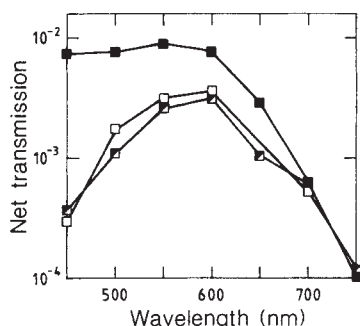


Fig. 3 Total transmission T plotted against wavelength at all three experimental sites illustrating the differences between the sites A, $\square$; B, $\blacksquare$; C, $\blacksquare$.

source²⁻⁴, a choice which limits the utility of the results in elucidating the roles of ice inhomogeneity and anisotropy in the light transport problem. Our approach was to place a collimated monochromatic light source on the surface of the ice, and measure the transmission by submerging detectors on a mechanical arm through a hole in the ice (Fig. 1). A chopped light source and a phase-lock detector separated our signal from background sunlight. The important new aspect of the technique is that it permits the measurement of the transmitted intensity as a function of the horizontal separation (ρ) of the source and detector⁵. The magnitude, spatial extent, and profile of the diffuse transmission yield information about the depth dependence of the scattering⁶. Finally, the transmission results were complemented by using a downward-viewing detector to measure the backscattered light escaping from the top surface of the ice, and its fall-off with source-detector separation.

We used the instrument in November 1985 on 1.6-m thick first-year sea ice at three sites (A, B, C) near Tent Island in McMurdo Sound. Site C was snow-free, but sites A and B had snow covers of 25 and 15 mm, respectively, which were cleared before taking the measurements reported here. Typical transmission data are shown in Fig. 2. Note that the light diffuses out to a horizontal width comparable to the thickness of the ice. The data for 700 nm radiation show a reduced intensity and width relative to the 500 nm data; both of these arise from the stronger absorption of 700 nm light in the bulk of the sea ice⁵.

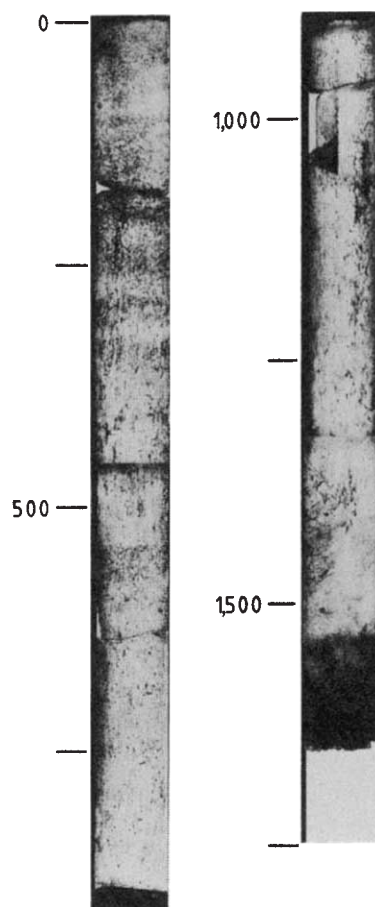


Fig. 4 Vertical core of sea ice from site B. Clearly seen is the granular surface layer (upper left) that gives way to the columnar structure in the bulk. The dark band at the base of the ice (lower right) is due to algae. The numbers down the left side of the core give the depth in mm.

Data like those in Fig. 2 can be integrated to give the net transmission results shown in Fig. 3. The three sites showed similar transmission reductions in the red, but at two sites (A and B) there was also a reduction in the blue, near 500 nm. This absorption in the blue was not accompanied by the narrowing of the under-ice profile mentioned above, a result which indicates that it did not take place through the bulk of the ice. We identify the blue absorption as occurring in a thin algal layer near the bottom of the ice⁷, which was found from ice cores to be more heavily populated at sites A and B than at site C.

We have interpreted our results with the aid of Monte Carlo simulations based on simple models for the scattering and absorption of light^{5,6,8-10}. The backscattered and transmitted profiles could not both be fitted by any model containing a uniform density of scattering centres within the ice. A satisfactory fit is obtained only by assuming a strongly scattering top layer lying above a bulk layer in which the scattering is anisotropic; both of these features are suggested by the graded ice structure¹¹ shown in the ice-core photograph of Fig. 4. In Fig. 2 we have shown a comparison between the measured profiles and those predicted by a model containing a 0.5-m thick top layer with an isotropic scattering length of 0.06 m, followed by ice in which a vertically travelling photon has an effective scattering length of 0.2 m, while for a photon travelling horizontally it is only 0.07 m. The anisotropic nature of sea ice is also observed in other physical properties such as the electrical conductivity¹². Absorption coefficients in the ice, treated as fitting parameters, are very close to the values given in the literature for pure ice¹³.

An interesting observation, based both on comparisons between submerged detectors with differing angular response and on direct angular measurements, was that the radiation emerging from the lower surface of the ice was directed primarily within about 30° of the vertical direction. This effect is a result of the anisotropic scattering in the lower layer; the Monte Carlo simulation^{5,6} with anisotropic scattering showed a similar strongly peaked angular distribution.

The scattering lengths (δ) determined by our measurements are of course related to the typical separation (l) between scattering surfaces, and for the predominantly forward scattering¹⁴ of air bubbles and brine pockets in ice,

$$\frac{l}{\delta} \approx (1 - \cos \theta)$$

where θ is the average angle through which light is scattered by one defect. This angle depends on the shape of the defect and on its refractive index. It is particularly small for pockets of brine, which have a refractive index very close to that of the surrounding ice. The measured scattering angles in NaCl ice¹⁵, combined with our scattering lengths, suggest a separation (l) of 3–6 mm, in agreement with the brine channel separations shown in Fig. 4. The strong scattering near the surface is due to the presence of air bubbles, particularly near the top of the granular layer.

Our data demonstrate that light scattering in sea ice shows strong inhomogeneity (layering) and anisotropy, and that these determine the transmitted radiation field. Near the surface the ice is relatively strongly scattering, and this top layer scatters most of the light back through the surface; the albedo can be expected to be a strong function of the character and depth of this layer. The bulk of the ice is made up of columnar ice, and

the anisotropic scattering in this material reduces the horizontal travel of the transmitted radiation, acting as a weakly channeling medium. It further imposes an angular dependence to the radiation entering the water, peaking it in the downward direction. An interesting consequence that can be inferred from the result is that the light throughout the bulk of the ice is strongly peaked in the vertical directions, and therefore, so is the light falling on the algae within the ice. It should be emphasized that this peaked distribution is a result of scattering by oriented brine channels, and is not related to the direction along which light is introduced into the ice. Sunlight, falling on the ice at relatively low angles, will still propagate through most of the ice with a radiance that is strongly peaked in the vertical directions.

We thank Simon Brown and Pat Langhorne for assistance with the Monte Carlo simulations and the ice-core photography, and Arnold Heine for general field assistance. The logistic support of the Antarctic Division of the New Zealand DSIR is also gratefully acknowledged.

Received 27 October 1986; accepted 3rd February 1987.

1. Garrison, D. L., Sullivan, C. W. & Ackley, S. F. *Bioscience* **36**, 243–250 (1986).
2. Weller, G. E. *Nature* **221**, 355–356 (1969).
3. Roulet, R. R., Maykut, G. A. & Grenfell, T. C. *Appl. Opt.* **13**, 1652–1659 (1974).
4. Maykut, G. A. & Grenfell, T. C. *Limnol. Oceanogr.* **20**, 554–563 (1975).
5. Buckley, R. G., Trodahl, H. J. & Langhorne, P. J. *Phys. Engng Lab. Report No.* 915 (1986).
6. Brown, S. A., Buckley, R. G. & Trodahl, H. J. *Phys. Engng Lab. Report No.* 948 (1986).
7. Sullivan, C. W. *et al. Antarct. J. U.S.* **18**, 177–179 (1983).
8. Ishimaru, A., Kuga, Y., Cheung, R. L.-T. & Shimizu, K. *J. opt. Soc. Am.* **73**, 131–136 (1983).
9. Groenhuis, R. A. J., Ferwerda, H. A. & Ten Bosch, J. J. *Appl. Opt.* **22**, 2456–2462 (1983).
10. Groenhuis, R. A. J., Ten Bosch, J. J. & Ferwerda, H. A. *Appl. Opt.* **22**, 2463–2467 (1983).
11. Langhorne, P. J. & Robinson, W. H. *Cold Reg. Sci. Technol.* **12**, 197–214 (1986).
12. Buckley, R. G., Staines, M. P. & Robinson, W. H. *Cold Reg. Sci. Technol.* **12**, 285–290 (1986).
13. Grenfell, T. C. & Perovich, D. K. *J. geophys. Res.* **86**, 7447–7450 (1981).
14. van de Hulst, H. C. *Multiple Light Scattering* Vol. 1 (Academic, New York, 1980).
15. Grenfell, T. C. & Hedrick, D. *Cold Reg. Sci. Technol.* **8**, 119–127 (1983).

Confirmation of Carboniferous clastic swamp communities

Robert A. Gastaldo

Department of Geology, Auburn University,
Alabama 26849-3501, USA

Tropical lowlands dominated by arborescent lycophytes are the principal foci for our understanding of Carboniferous peat and non-peat accumulating coastal swamps^{1,2}. Suggested palaeoecological interpretations of various lycophytes and accessory vegetation have been based on transported compression assemblages^{3,4} and palynomorphs^{5,6}, or *in situ* permineralized peats⁷. Where transported macrodetritus, resulting in compression assemblages, is used to reconstruct past vegetational patterns, taphonomic factors tend to bias the preserved flora^{8–11}. Taken alone, or out of depositional context, the heterogeneity of an assemblage most probably bears little or no relationship to that of the original contributing communities. Gastaldo⁹ suggests that reconstructions should be based on the autecology of erect *in situ* plants because, with peats permineralized *in situ*, precise resolution in time of laterally equivalent vegetation is difficult. Here I document an erect *in situ* Early Pennsylvanian lycophyte swamp which provides the first confirmatory evidence for previously hypothetical ecological amplitudes.

The Early Pennsylvanian in northeastern Alabama (Westphalian A equivalent)¹² represents a southwestwardly prograding delta with aerial swamps and subaqueous estuarine sediments^{1,13}. Contour mining has exposed the sequence over ~45 km². Along Berry Mountain (Fig. 1), an en echelon stacked distributary system is flanked by mudstones and shales. The channel migrated laterally¹, breached a levee-bound swamp,

and deposited large volumes of splay sands and alluvial muds. One identified crevasse (SE1/4 SW1/4 T11S R2E, sec 31 Fig. 1) was probably the principal site of influx. A facies transition between channel sands and silty mudstones of the levee-back levee exists (NW1/4 SE1/4 T11S R1E Sec36).

Erect *in situ* biomass at this site, cast similarly to the 'Fossil Grove', Victoria Park, Glasgow¹⁴, represents three form-genera, *Lepidophloios*, *Lepidodendron* and *Sigillaria*. Basal diameters of mature trees exceed 1m; preserved heights attain 3 m (ref. 1). Identifiable leaf cushions are dependent on the maturity of the tree and its taxonomic affinity. Supplementary data are available from three litter beds within the 3-m moulding sequence. These represent forest floor accumulation of aerial organs after death, and macrodetritus transported minimally during flooding^{1,9}. Forest floors are distinguished by infilled prostrate trunks similar to those in Holocene analogues¹⁵. Phytological components of forestation were investigated along a transect from the channel sandstone—mudstone facies contact southeastward into mudstone exclusive lithologies (Fig. 1).

Standing biomass form-taxa include *Lepidophloios laricinus* Sternberg, *Lepidodendron aculeatum* Sternberg and *Sigillaria cf. elegans* (Sternberg) Brongniart. A distinctive gradient is observed near the channel sands. *Lepidophloios*, moulded and cast in mudstone, are found exclusively in distal sites (Fig. 1; SE1/4 NE1/4 T12S R2E Sec 6). Unidentifiable erect lycophytes are believed to be *Lepidophloios*. Only compression trunk and crown macrodetritus is found in this forest floor litter; the remaining 1% is calamitalean. *Lepidodendron* trees appear intermixed with *Lepidophloios* closer to the channel—levee subfacies (SW1/4 SE1/4 T11S R2E Sec 31). Trees are moulded and cast in mudstones or sandstones according to their position adjacent to the splay sands. Lycophyte densities are documented elsewhere¹. No evidence of erect *Sigillaria* occurs here; unidentifiable trees are either *Lepidophloios* or *Lepidodendron*. Trunks and crown branches (including strobili) of these form-genera

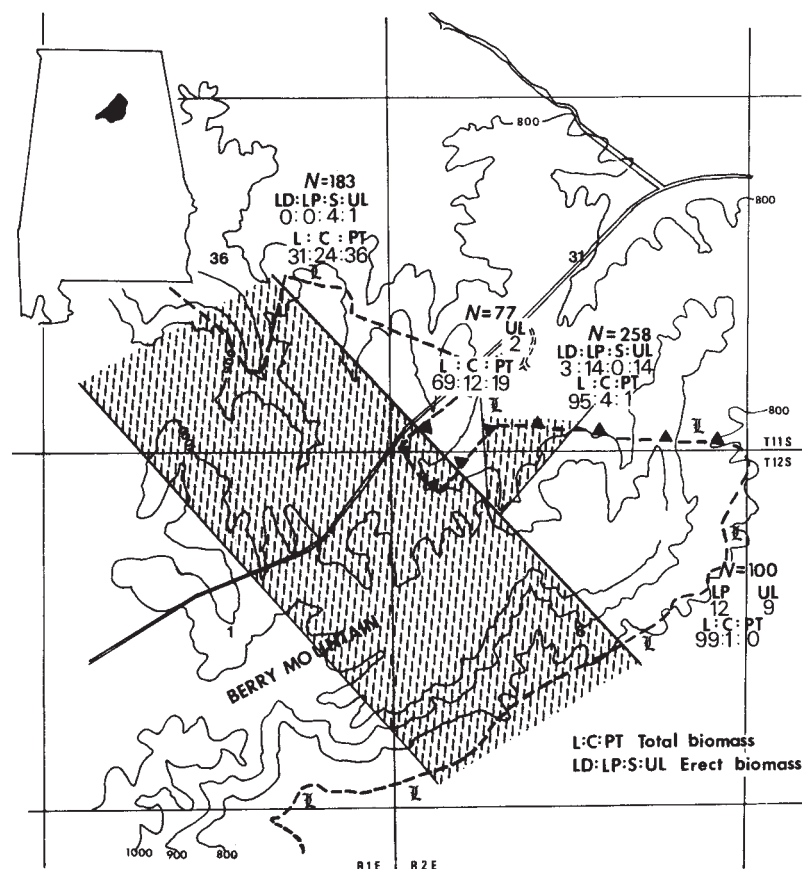


Fig. 1 Location of study area, Blunt County, in Alabama. Large map, Clarence, Alabama, 7 1/2' quadrangle quarter-quarter subdivisions of sections (sec.) are referred to in text in appropriate township (T) and range (R) coordinates. Stipple, en echelon-stacked channel system and crevasse splay; dashed line, contour mine highwall; Σ , *in situ* lycopod occurrences. Quantification of erect taxa presented by number of individuals in highwall at time of survey (LD, *Lepidodendron*; LP, *Lepidophlois*; S, *Sigillaria*; UL, unidentified lycopod trunk). Quantification of coalified compression assemblages of basal forest floor litter presented as percentages of major groups (L, lycophytes, C, sphenopsids, PT, pteridophytes and pteridosperms) based on a modification of the technique of Pfefferkorn *et al.*²². N, number of individuals used in determining percentages.

are in the forest floor litters. The site closest to the channel sandstones (NW1/4 SW1/4 T11S R1E Sec 36) preserve only identifiable *Sigillaria*. The observed number is low, probably because of excavation¹. Unidentifiable lycophytes are uncommon, but because rare compression *Lepidodendron* and *Lepidophlois* branches occur, it is not possible to distinguish a coexisting taxon, if any. *Lepidophlois* strobili and lateral branches from which strobili were abscised are more common. This debris may have been redistributed during flooding.

A second biomass trend is documented in the compression assemblages. The preserved forest floor litter provides evidence for increasing diversity towards the channel. An internal branch pith cast of *Calamites* Brongniart is the only accessory vegetation in *Lepidophlois*-exclusive sites (Fig. 1). The disarticulated specimen was probably redistributed during flooding. Its state of preservation contrasts markedly with *Calamites* proximal to the channel.

Assemblages from *Lepidodendron*-inhabited sites are taxo-

nomically more heterogeneous; the proportion of lycophytes decreases as effects of the splay are encountered. The closer the site is to the crevasse splay, the more numerous are the *Calamites* pith casts and pteridosperm elements. Pith casts may dominate sandstone lags to the exclusion of other plant organs. Pteridosperm elements adjacent to the splay include: disarticulated pinnules of arborescent medullosans *Neuropteris pocahontas* White and *N. smithii* Lesquereux; small fragments of medullosan stems; and pinnae of the lyginopterid pteridosperm, *Lyginopteris hoeninghausii* (Brongniart) Gropp. Ultimate pinnae of *Sphenopteris pottsvillea* (White) Gastaldo and Boersma are rare. No evidence of ferns (pinnae, disarticulated fronds or rachis, stems) occurs here unless recovered telomic systems of small dichotomizing branches are pteridophyte remains.

Silty mudstone compressions proximal to the channel sands provide a unique insight into understory composition and alluvial swamp diversity. Attached or isolated leaves and rare strobili make up the lycophyte contribution. Fragmented pros-

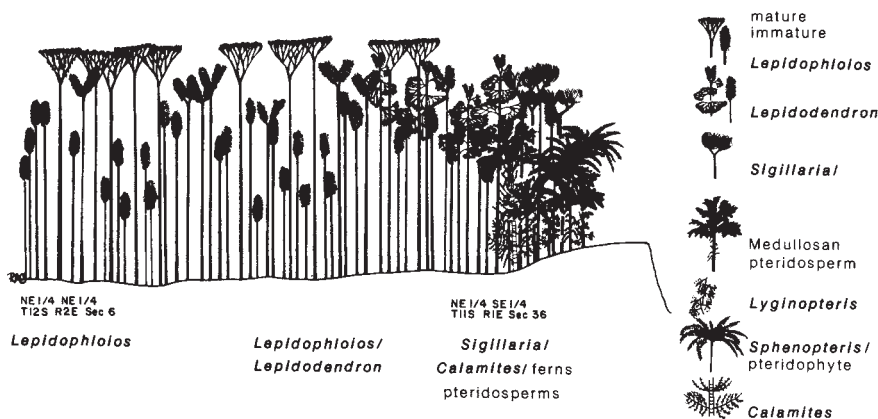


Fig. 2 Reconstruction of transect through Early Pennsylvanian levee-bound deep swamp to back-levee based on distribution of erect and prostrate trees and lianas. Spacing of immature and mature trees based on published lycophyte distribution¹, although lateral distance is not to scale. Growth habit of trees based on reconstructions by Wnuk²³, DiMichele²⁴, Hirmer¹⁷, and Pfefferkorn *et al.*²⁵.

trate medullosan pteridosperm stems, with attached rachis, exceed 2.5 m in length. Their configuration is similar to specimens ascribed to a tree/shrub habit^{16,17}. Large frond fragments, broken through excavation, may be larger than 1 m. Systematic affinities include *N. pochontas*, *N. smithii* and *S. pottsvillea*. Entire lyginopterid fronds are attached to metres of coiled stem (3 cm diameter). These are interpreted as lianas in growth habit probably coiling upon themselves during their fall to the forest floor. Pteridophytes are represented by entire tree apices (presumably with trunks attached before mining) with attached rachis². Associated *Sphenopteris* foliage may be correlative. *Calamites* stems with attached lateral branches and foliage of *Asterophyllites* display external morphology. No evidence exists for stem/branch degradation common in this taxon. No sphenophyllaleans have been recovered.

The prostrate organs preserved on the lowest bedding plane represent the forest floor litter at the time of levee breach and indicate that macrodetritus was transported minimally within the levee-bound swamp. Storm-induced, traumatically generated assemblages sample only canopy organs, principally leaves^{11-13,19}. In Holocene analogues, woody materials from levee and back-levee sites often are redistributed into the periphery of alluvial swamps during major flooding^{10,15}. Floating stumps, trunks, and large branches may become wedged between erect vegetation such as trees, so although they may remain afloat until flooding subsides, such organs rarely travel significant distances within the swamp. The assemblage of medullosan trunks, lyginopterid interpreted lianas, pteridophyte crowns, and nearly complete fronds characterize a minimally transported assemblage. These plants are interpreted as components of the forest floor litter proximal to the levee, and were not transported into the deep swamp by breaching floodwaters. If they had been transported by the initial breach, a similar assemblage should be preserved proximal to the breach site, but this has not been observed. As flood waters filled the levee-bound swamp, this litter probably was introduced into the water column, was possibly minimally redistributed, and settled near the growth site. Suspension load sedimentation covered the macrodetritus, and standing water prevented oxidation. Forest litter throughout the swamp was preserved similarly. Subsequent litter horizons were generated by the death of the standing lycophytes. Trunks not buried in basal sediments rotted, fell, and crown organs abscised. Death and crown breakup were not simultaneous throughout the forest as evidenced by two supradjacent forest litters dominated by lycophytes.

This breach of the levee, preserving as it did erect *in situ* lycophytes and minimally transported laterally equivalent riparian vegetation, has provided the first documented evidence of Carboniferous swamp ecological amplitudes. Previous reconstructions^{3,4} relied on hypotheses generated from either temporally segregated samples, such as permineralized peats, or geographically isolated fragments in which lateral equivalents are unknown, such as most coalified compression assemblages.

Lepidophloios was the most water-stress tolerant taxon in pure stands distal to the channel. In Holocene levee-bound swamps standing water is present in distal sites during some part of the year, often for extended periods. Very low diversity characterizes these saturated edaphic conditions. Where microtopographic relief may have provided for slightly better-drained alluvial muds, diversity increases with identifiable *Lepidodendron*. At locations closest to the channel, where topographic relief is greatest (levees and back-levees), a change in lycophyte vegetation occurs. *Sigillaria* dominates to the apparent exclusion of *Lepidophloios*. Here biodiversity increases with the introduction of understory vegetation. The proximity of the site to the channel suggests that soils were better drained providing edaphic conditions capable of supporting a lush vegetation. Similar patterns are documented in Holocene analogues^{8,20}. Based on the spatial relationships in the *Lepidophloios* and *Lepidodendron*-

dendron dominated sites¹, vegetational communities are reconstructed from interior swamp to back-levee (Fig. 2).

This window in time confirms suggested hypothetical amplitudes of swamp-dwelling lycophytes^{7,21} where *Lepidophloios* is considered an inhabitant of wetter habitats, *Lepidodendron* a plant of various intermediate tolerance, and *Sigillaria* indicative of drier sites. The lateral spatial relationships of this Early Pennsylvanian alluvial swamp document the first account of a lycophyte ecotone encompassing understory vegetation including pteridosperm, pteridophyte and sphenopsid taxa. Biodiversity distribution of Early Pennsylvanian alluvial swamps is similar to that of Holocene swamps. Low diversity characterizes distal swamp communities with an accompanying increase with slight changes in topography associated with overbank levee deposition.

Received 28 November 1986; accepted 16 February 1987.

1. Gastaldo, R. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **53**, 191-212 (1986).
2. Phillips, T. L., Peppers, R. A. & DiMichele, W. A. *Int. J. Coal Geol.* **5**, 43-109 (1985).
3. Scott, A. C. *Palaeontology* **20**, 447-473 (1977).
4. Scott, A. C. *Proc. Yorks. geol. Soc.* **41**, 461-508 (1978).
5. Smith, A. H. V. *Geol. Mag.* **94**, 345-363 (1957).
6. Smith, A. H. V. *Proc. Yorks. geol. Soc.* **33**, 423-474 (1962).
7. Phillips, T. L. & Peppers, R. A. *Int. J. Coal Geol.* **3**, 205-255 (1984).
8. Scheiching, M. H. & Pfefferkorn, H. W. *Rev. Palaeobot. Palynol.* **41**, 205-240 (1984).
9. Gastaldo, R. A. C. *R. Dixième Congr. int. Strat. Geol. Carb.* **2**, 281-296 (1985).
10. Gastaldo, R. A. *Ala. Geol. Surv. Reprint. Ser.* **66**, 1-35 (1985).
11. Spicer, R. A. in *Biostratigraphy of Fossil Plants* (eds Dilcher, D. & Taylor, T. N.) 171-183 (1980).
12. Gastaldo, R. A. *Third N. Am. Paleon. Conf. Proc.* **1**, 187-192 (1982).
13. Gibson, M. A. & Gastaldo, R. A. *J. Paleont.* (in the press).
14. Gastaldo, R. A. *Scott. J. geol.* **22**, 77-83 (1986).
15. Gastaldo, R. A. *Univ. Tenn. Dept. geol. Sci. Stud. Geol.* **15**, 27-44 (1986).
16. Wnuk, C. & Pfefferkorn, H. W. *Rev. Palaeobot. Palynol.* **41**, 329-351 (1984).
17. Hirmer, M. *Handbuch der Paläobotanik* (Oldenbourg, München, 1927).
18. Scheiching, M. H. *Argum. Paleobot.* **6**, 133-138 (1980).
19. Ferguson, D. K. *Rev. Paleobot. Palynol.* **46**, 117-188 (1985).
20. Stout, J. P., Lelong, M. J., Dowling, H. M. & Powers, M. T. *Ala. Coastal Area Bd. Tech. Rep.* **81-49A**, 1-25 (1982).
21. DiMichele, W. A. & Phillips, T. L. *Rev. Palaeobot. Palynol.* **44**, 1-26 (1985).
22. Pfefferkorn, H. W., Mustafa, H. & Hass, H. *Neues Jb. Geol. Paläont. Abh.* **150**, 253-269 (1975).
23. Wnuk, C. *Palaeontographica* **195B**, 153-181 (1985).
24. DiMichele, W. A. *Palaeontographica* **171B**, 57-77 (1979).
25. Pfefferkorn, H. W., Gillespie, W. H., Resnik, D. A. & Scheiching, M. H. *Mosasaur* **2**, 1-8 (1985).

The origin of egg-laying mammals

Zofia Kielan-Jaworowska*, A. W. Crompton† & Farish A. Jenkins, Jr†

* Zakład Paleobiologii, Polska Akademia Nauk, al. Zwirki i Wigury 93, 02-089 Warszawa, Poland
† Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

The origin of monotremes—egg-laying mammals from the Australian region—and their relationship to therians (marsupials and placentals) have been intriguing zoological problems. Recently Archer *et al.*¹ described a fragment of a lower jaw of an early Cretaceous monotreme from Australia (*Steropodon galmani*) and interpreted its teeth as tribosphenic and thus supposedly derived from a relatively advanced stage in therian evolution. This discovery is of major importance, and radically alters currently held opinion that monotremes and therians diverged at the earliest stage of mammalian evolution. We agree that *Steropodon* is a therian but would argue that its teeth are not tribosphenic. *Steropodon* appears to have been derived from therians before the development of tribosphenic teeth, possibly during the Jurassic period.

The earliest known tribosphenic molars are of early Cretaceous age²⁻⁴ and represent a lineage leading from the eupantotheres to the Tribosphenida^{5,6} (a legion encompassing the Eutheria, Marsupialia and several poorly known Cretaceous groups: Aegialodontia, Pappotherida⁷ and Deltatheroidea⁸). An essential characteristic of the tribosphenic pattern is the develop-

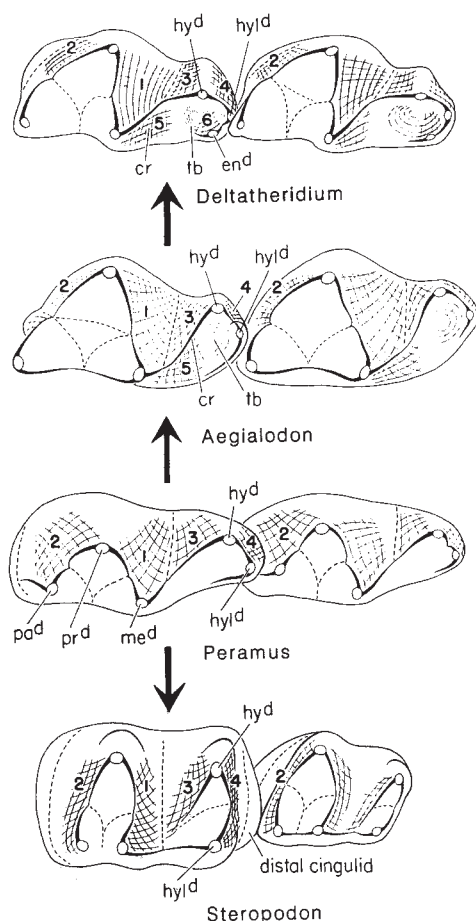


Fig. 1 Comparison of pre-tribosphenic (*Peramus*), tribosphenic (*Aegialodon* and *Deltatheridium*) and *Steropodon* lower right molars in occlusal view. Homologous shearing surfaces (1-6) are numbered according to the scheme of Crompton³. cr, Cristid obliqua; en^d, entoconid; hy^d, hypoconid; hyl^d, hypoconulid; me^d, metaconid; pa^d, paraconid; pr^d, protoconid; tb, talonid basin. The buccal surface is toward the top of the figure; mesial is to the left.

ment of an additional lingual cusp (protocone) that occludes in a talonid basin of the lower molar (Fig. 1). Typically the talonid basin is surrounded by three cusps (hypoconid, hypoconulid, entoconid); the entoconid is a feature that arose in specific relation to a shearing plane with the protocone.

In *Steropodon* (cast kindly supplied by Dr Alex Ritchie) the talonid is greatly enlarged but in a different manner from that of tribosphenic molars (Fig. 1). There are two talonid cusps that are nearly as high as those of the trigonid with only a small basin between them. The basin does not appear to have functioned as a major occlusal surface (as it does in tribosphenic molars).

In *Steropodon*, two major areas of shear are developed. The first, between the distal surface of the trigonid and mesial surface of the hypoconid (facets 1 and 3, Fig. 1), is a greatly enlarged embrasure (hypoflexid) for the paracone. The cristid obliqua does not appear to extend up the distal surface of the metaconid, so that a single shearing plane is present on the distal surface of the trigonid. In tribosphenic molars the cristid obliqua is a characteristic feature of the talonid that passes from the hypoconid up the posterior surface of the metaconid, and demarcates occlusal areas for the protocone (lingually) and paracone (buccally). This difference strengthens the interpretation that a protocone was not developed in *Steropodon*.

The second area of shear is developed on the distal surface of the talonid and mesial surface of the trigonid of the adjacent tooth (facets 4 and 2, Fig. 1). Although wear on the posterobuccal

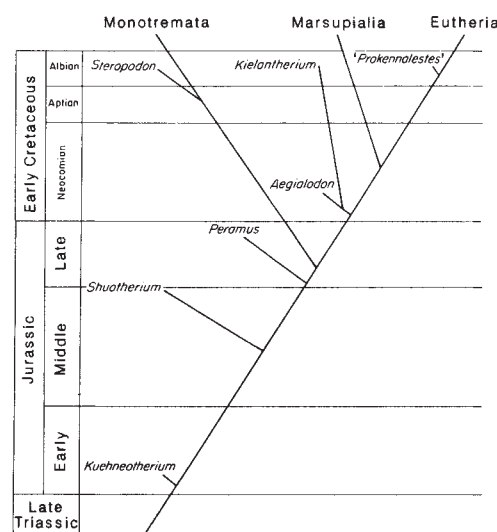


Fig. 2 Relations of *Steropodon*.

surface of the talonid (by the metacone) is a common feature in early Theria, this shearing plane appears to have been greatly enlarged in *Steropodon*. Comparing this facet with that of early Theria, we interpret the two talonid cusps as a hypoconid and hypoconulid. Archer *et al.*¹ interpreted the latter cusp as an entoconid; we regard this as unlikely because its buccal surface lacks wear (from a protocone) but supports a major shearing plane distally, which the entoconid never does in tribosphenic molars.

In tribosphenic molars and in the molars of advanced eupantotheres such as *Peramus*⁹, the hypoconulid contacts the next tooth. The triangular embrasure between teeth is formed by the hypoconid, hypoconulid and mesial surface of the trigonid; if an entoconid is present, it is excluded from this embrasure. The unusual enlargement and lingual extension of this embrasure in *Steropodon* has resulted in intermolar contact by mesial and distal cingulids that lie below the shearing surfaces. Although the upper molars of *Steropodon* are as yet unknown, the development of this narrow, transversely elongate shearing embrasure between adjacent lower molars probably entailed a transverse, loph-like arrangement of the metacone and metastyle.

On the basis of their interpretation of *Steropodon* dental structure, Archer *et al.*¹ suggested that *Steropodon* represents an evolutionary stage (albeit specialized) between *Kielantherium* and *Aegialodon*. In fact, *Kielantherium* is extremely similar to *Aegialodon*. One possible difference is the presence of a third cusp flanking the talonid basin in *Aegialodon*. However, Kermack *et al.*² referred to this cusp as an entoconid? because it is so poorly preserved; the existence of a cusp in that region is in fact uncertain. Another difference is that the talonid basin is slightly wider in *Aegialodon*. These differences appear so trivial that *Kielantherium* has been considered a junior synonym of *Aegialodon*¹⁰. Although we regard *Kielantherium* as a separate genus, we believe that its similarity to *Aegialodon* justifies the assignment of these taxa to the same family.

On the other hand, *Steropodon* differs so dramatically from both *Aegialodon* and *Kielantherium* (two very high talonid cusps, lack of a true talonid basin, hyperdevelopment of the shearing embrasures, large horizontal anterior, buccal and posterior cingulids, and the apparent lack of the protocone), that its teeth cannot be regarded as tribosphenic. Archer *et al.*¹ do exclude *Steropodon* from the Tribosphenida in their cladograms, but their interpretation of the lower molar cusp pattern implies a closer relationship between *Steropodon* and tribosphenids than we believe is warranted.

The elaboration of a large occlusal area behind the talonid in *Steropodon* would at first seem extraordinary for a Mesozoic

therian. However, there is now solid evidence that at least one other lineage of Jurassic mammals developed complex, but non-tribosphenic, occlusion from the reversed-triangle pattern of cusps. *Shuotherium dongi*¹¹, from the late Jurassic of China, possesses a 'talonid' anterior to the trigon. As in the case of *Steropodon*, the specialization of the upper molars would seem to have been in the area of the metacone, but because there is a true 'talonid' basin in *Shuotherium* a 'pseudoprotcone' may have been present. The molar pattern of *Steropodon* was possibly derived from *Peramus*-like forms (Figs 1, 2). *Peramus* has a talonid with a hypoconid and a hypoconulid, but lacks a protocone and hence also a basin in the talonid. Modifications of this molar type toward the *Steropodon* pattern would have involved a transverse widening of the shearing surface distal to the hypoconulid and hypoconid, enlargement of these two cusps, and enlargement of the hypoflexid to accommodate the paracone.

*Aegialodon*² from the Wealden (Valanginian) is the earliest known tribosphenic molar. It is therefore reasonable to conclude

that the common ancestor of *Aegialodon* and *Steropodon* must have existed in Jurassic times. We agree with Kemp¹² that monotremes were derived from early therians, but conclude that the dichotomy occurred well before the origin of the Tribosphenida.

Received 18 March 1986; accepted 11 February 1987.

1. Archer, M., Flannery, T. F., Ritchie, A. & Molnar, R. E. *Nature* **318**, 363–366 (1985).
2. Kermack, K. A., Lees, P. M. & Mussett, F. *Proc. R. Soc. B* **162**, 535–554 (1965).
3. Crompton, A. W. in *Early Mammals* (eds Kermack, D. M. & Kermack, K. A.) 65–87, *Zool. J. Linn. Soc.* **50**, Suppl. 1 (1971).
4. Crompton, A. W. & Kielan-Jaworowska, Z. in *Development, Function and Evolution of Teeth* (eds Butler, P. M. & Joysey, K. A.) 249–287 (Academic, London, 1978).
5. McKenna, M. C. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
6. Dashzeveg, D. & Kielan-Jaworowska, Z. *Zool. J. Linn. Soc.* **82**, 217–227 (1984).
7. Butler, P. M. *Breviora* **446**, 1–27 (1978).
8. Kielan-Jaworowska, Z. *Palaeont. Pol.* **46**, 173–183 (1984).
9. Clemens, W. A. & Mills, J. R. E. *Bull. Br. Mus. nat. Hist. Geol.* **20**, 87–113 (1971).
10. Fox, R. C. *Can. J. earth Sci.* **13**, 1105–1118 (1976).
11. Chow, M. & Rich, T. H. V. *Aust. Mammal.* **5**, 127–142 (1982).
12. Kemp, T. S. *Zool. J. Linn. Soc.* **77**, 353–384 (1983).

Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision

Patrick J. Bennett* & Martin S. Banks†

* Department of Psychology, and † School of Optometry and Department of Psychology, University of California, Berkeley, California 94720, USA

The ability to detect, discriminate and identify spatial stimuli is much poorer in the peripheral than in the central visual field. Some deficits are eliminated by scaling stimulus size. For example, grating detectability is roughly constant across the visual field when spatial frequency and target extent are scaled appropriately^{1,2}. Other deficits persist despite scaling. For instance, some readily detectable patterns are more difficult to identify peripherally than in the fovea^{3,4}. This deficit is caused, at least partially, by a reduced ability to encode spatial phase (or relative position)^{5–7}. To specify the properties of foveal and peripheral phase-encoding mechanisms, we measured discrimination thresholds for compound gratings at several eccentricities. Our observations are consistent with a two-channel model of phase encoding based on even- and odd-symmetric mechanisms⁸ (see Fig. 1), but the sensitivity of the odd-symmetric mechanisms decreases dramatically with eccentricity. Thus, the loss of sensitivity in one type of mechanism may underlie the reduced ability to encode spatial phase peripherally.

We used a paradigm introduced by Field and Nachmias⁸. Observers were presented with two compound gratings composed of a fundamental (F) and second harmonic (2F) added in two phases differing by 180°. A 180° phase shift does not alter the peak-to-trough amplitude, or the contrast, of these waveforms (Fig. 1). Discrimination thresholds for several 180° phase shifts (for example, 0–180, 45–225) were measured by varying the contrast of 2F. These thresholds can be explained by the operation of two broad-band channels. One channel has positive and negative even-symmetric receptive fields, and the other has positive and negative odd-symmetric fields (Fig. 1). Only mechanisms located at particular points in the waveform contribute to phase discrimination: even-symmetric mechanisms centred on peaks and troughs of F and odd-symmetric mechanisms centred on zero-crossings of F. This assumption differs from the Field and Nachmias model⁸; they assumed that the computations were restricted to these points in the waveform when the contrast of 2F was low relative to the contrast of F. We assumed that the computations are spatially restricted in

this way for all 2F contrasts. As it turns out, this assumption is required to explain our results in the far periphery, a point we expand below.

Figure 1 illustrates how the model discriminates 180° phase shifts (for explanation see Fig. 1 legend). Two important features of the model need to be emphasized. First, a 0–180 (degree) phase shift does not change the responses of odd-symmetric mechanisms centred on zero-crossings in F. Second, a 90–270 phase shift does not change the responses of even-symmetric mechanisms centred on the peaks and troughs of F. Thus, discrimination thresholds for the 0–180 and 90–270 conditions reveal the sensitivities of the even and odd channels, respectively. From these thresholds, the two-channel model predicts thresholds for all other 180° phase shifts (Fig. 2). We hypothesized that the peripheral phase-encoding deficit is caused by the reduced sensitivity of one or both of these mechanisms.

The stimuli were generated on a Joyce Electronics display with P4 phosphor. The targets were presented in a 4.5 × 15 cm window surrounded by a dark field. Target size was held constant and viewing distance was varied to test different retinal eccentricities. This procedure allowed us to scale target size appropriately for each eccentricity. The gratings were vertically oriented and contrast was modulated by a gaussian envelope with a standard deviation of 1.5 cm. Space-average luminance was 95 cd m⁻². A light-emitting diode served as the fixation target when stimuli were presented peripherally. Observers fixated the centre of the window for foveal conditions.

Before the discrimination experiments, contrast sensitivity functions (CSFs) were measured at the fovea and at 5°, 10°, 20° and 40° in the temporal retina. In the discrimination experiments, the spatial frequency of F was set to the peak of the CSF, and its contrast was set about 1 log unit above detection threshold. Thus, F was equally visible at all eccentricities. In some experimental sessions the contrast of 2F was varied according to 2-down/1-up staircase rule. The method of constant stimuli was used in other sessions. Trials were started by the observer and consisted of two 250 ms intervals. Observers judged whether the patterns in each interval were the same or differed by a 180° phase shift of 2F relative to F. Auditory feedback indicated the accuracy of the response after each trial.

Two potential problems arise when using this procedure. First, a 180 deg phase shift changes the average luminance of a gaussian-damped grating. To ensure that the observers could not use average or local luminance cues to make discriminations, the waveform was displaced randomly within the gaussian envelope between intervals on each trial. Second, suprathreshold gratings are known to mask gratings of lower spatial frequency^{9–11}. Thus, raising the contrast of 2F to obtain discrimination thresholds may reduce the visibility of F. In all sets of

Fig. 1 Receptive fields of the mechanisms and the luminance profiles of the stimuli. Inset, positive and negative even- and odd-symmetric mechanisms. Except for the requirement that each mechanism responds to F and $2F$, the model's predictions are not dependent on the bandwidths of the mechanisms. The stimuli were pairs of compound gratings composed of a fundamental (F) and second harmonic ($2F$). The luminance profiles of the stimuli are shown with $2F$ added in 90° and 270° (*a*), 30° and 210° (*b*) and 0° and 180° (*c*) phase. The model uses changes in the response of mechanisms centred on the peaks, troughs and zero-crossings in F to discriminate phase shifts. To illustrate this process, we have represented the responses of even and odd mechanisms with light and dark arrows, respectively. Responses are computed by cross-correlating the stimulus with the receptive fields and half-wave rectifying; arrow length is proportional to response magnitude. Notice that a 0 – 180 (degree) shift (*c*) alters the responses of even-symmetric mechanisms, but not odd-symmetric mechanisms. This is illustrated by the lengths of the arrows in corresponding locations in the two waveforms. The length of light arrows pointing upwards (positive even-symmetric response) decreases when $2F$ is shifted from 0° to 180° phase, whereas the length of light arrows pointing downwards (negative odd-symmetric response) increases when $2F$ is shifted from 0° to 180° phase. The lengths of all dark arrows (positive and negative odd-symmetric response) are the same in both waveforms. A 0 – 180 shift is discriminated, therefore, when the change in responses of even-symmetric mechanisms exceeds some threshold. In a 90 – 270 phase shift (*a*), the reverse is true: responses of odd-symmetric mechanisms, but not even-symmetric mechanisms, are altered. Again, this is illustrated by the lengths of arrows in corresponding locations in the two waveforms. A 90 – 270 shift is discriminated when changes in odd-symmetric mechanisms exceed some threshold. A 30 – 210 phase shift changes the responses of even and odd-symmetric mechanisms (*b*). The two-channel model predicts that a 30 – 210 shift is discriminated when changes in either the even-symmetric or the odd-symmetric mechanisms reach threshold (see Fig. 2). The $2F$ contrasts needed to discriminate 0 – 180 , 30 – 210 and 90 – 270 phase shifts are plotted in Fig. 2 as M_0 , M_{30} , and M_{90} , respectively.

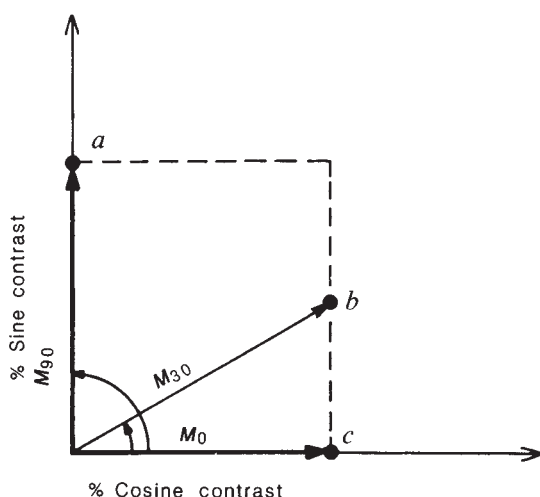
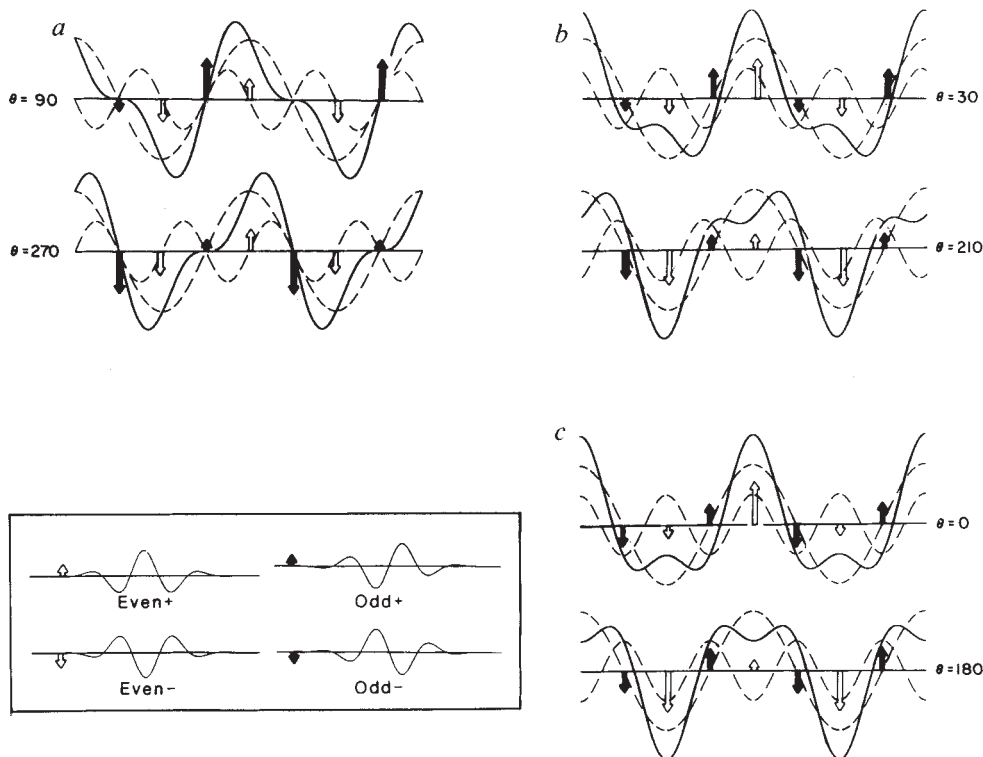


Fig. 2 Plot of predicted discrimination thresholds. The angle expansion formula $M \cos(2\pi 2F + \theta) = M \cos \theta \cos(2\pi 2F) - M \sin \theta \sin(2\pi 2F)$, is used to analyse $2F$ into its cosine and sine components. For any 180° shift, the predicted threshold is the contrast needed for the cosine amplitude to equal the 0 – 180 threshold, or for the sine amplitude to equal the 90 – 270 threshold. The predicted and observed thresholds are plotted as follows: the length of a vector is proportional to the contrast of $2F$, and phase is represented by its angle. The 0 – 180 threshold (*c*) is plotted on the abscissa and the 90 – 270 threshold (*a*) on the ordinate. The model predicts that an 180° shift is discriminated when the cosine or sine components of $2F$ equal one of those thresholds. Thus, the predicted thresholds lie along the two dotted lines, the horizontal line representing 180° shifts discriminated by odd-symmetric mechanisms, and the vertical line representing shifts discriminated by even-symmetric mechanisms. The prediction for a 30 – 210 shift (*b*) is illustrated.

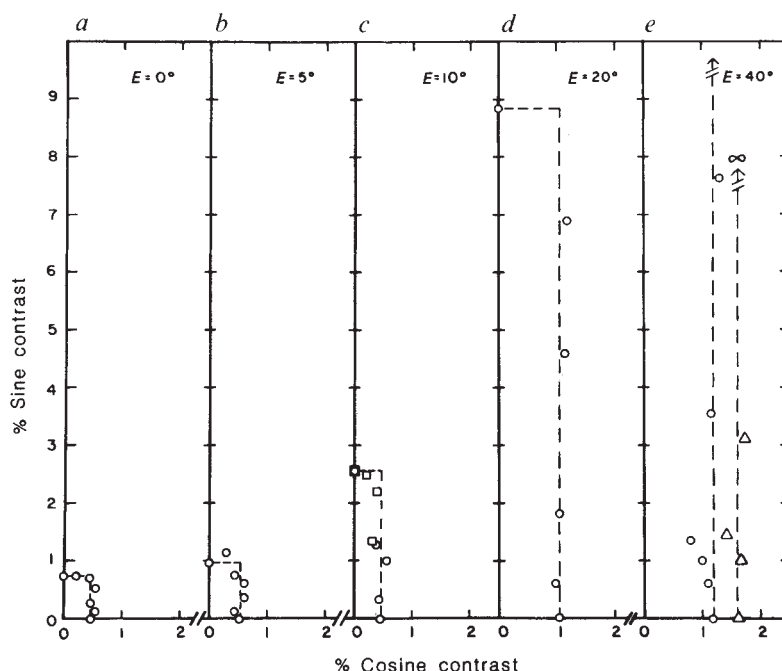
conditions reported here, however, F was clearly visible despite the presence of $2F$. A failure to discriminate phase shifts cannot, therefore, be attributed to an inability to detect both grating components.

Discrimination thresholds were similar in the fovea for the 0 – 180 and the 90 – 270 phase shifts (Fig. 3*a*), and thresholds for the other phase shifts were quite consistent with the predictions of the two-channel model. This implies that two mechanisms with similar sensitivities operate in the fovea, replicating the

results of Field and Nachmias⁸. With increasing eccentricity, the thresholds for the 0 – 180 shift rose slightly, but thresholds for the 90 – 270 shift rose monotonically and dramatically (Fig. 3*b–e*). At 20° , for example, the 90 – 270 threshold was ~ 10 times greater than the 0 – 180 threshold. Thus, two mechanisms also seem to operate in the peripheral visual field (with the possible exception of 40°), but their sensitivities are strikingly dissimilar.

The difficulty of the 90 – 270 discrimination in the periphery has an interesting implication for the model. Even-symmetric

Fig. 3 Phase discrimination thresholds at five eccentricities (E): *a*, 0° ; *b*, 5° ; *c*, 10° ; *d*, 20° ; *e*, 40° . The circles represent the thresholds from one observer using a 2-down/1-up staircase procedure. Thresholds obtained with the method of constant stimuli are represented by the squares in *c*. Thresholds from a second observer, obtained with the staircase procedure, are shown as triangles in *e*. When the staircase procedure was used, threshold was defined as the geometric mean of the final 12 reversals. When the method of constant stimuli was employed, threshold was the 70% point on the psychometric function defined by 4 points with at least 100 trials per point. Both observers had normal vision, and viewed the stimuli monocularly with natural pupils. At an eccentricity of 40° , a 90–270 phase shift could not be discriminated. The predictions of the two-channel model are illustrated by the dotted lines in *a–e*. Values of F , the fundamental spatial frequency (in cycles per degree; c per deg) were: for *a*, 4; *b*, 2; *c*, 1.5; *d*, 1; *e*, 0.5. C_F , contrast of the fundamental, was 10% throughout.



mechanisms could readily discriminate these two phases if they made their computations at points other than the peaks and troughs of F . One can demonstrate this by computing the responses of the mechanisms by cross-correlating the $F+2F$ waveform with an even-symmetric receptive field (Fig. 1). The difference between the responses to the 90° and 270° stimuli is a sinewave of frequency $2F$, and the response differences are greatest at points distant from the peaks and troughs of F . Even-symmetric mechanisms positioned at such points should, therefore, be able to make the 90–270 discrimination at high contrasts of $2F$. The fact that the 90–270 discrimination is impossible in the far periphery suggests that even-symmetric mechanisms make their computations only at or near the peaks and troughs of F where the differential response is zero.

The stimuli in this experiment were scaled to render them equally detectable at all eccentricities. The scaling factor was similar to previous estimates of the retinal magnification factor^{2,12}. Livingston and Hubel¹³ argued that a different scaling factor should be used in some phase discrimination tasks. We followed their suggestion by examining whether different scaling factors would reduce the disparities between foveal and peripheral performance. Discrimination thresholds were measured in the fovea and at 40° using a range of spatial frequencies. Again, the contrast of F was set about 1 log unit above detection threshold. We found that foveal thresholds for the 0–180 and 90–270 discriminations are similar to one another for all the spatial frequencies we tested (Fig. 4*a*). In the far periphery (Fig. 4*b*), thresholds for the 0–180 discrimination are similar at the two frequencies, and the 90–270 discrimination is impossible at both. Thus, the foveal-peripheral disparity is scale-invariant at least over the range of frequencies tested. We conclude that Livingston and Hubel's scaling hypothesis does not account for the differences we observe between foveal and peripheral phase discrimination.

Our findings are consistent with earlier reports of phase-encoding anomalies in the periphery^{6,7}, but they extend those reports by suggesting that the anomalies are caused primarily by sensitivity losses among odd-symmetric mechanisms. Consequently, our results are inconsistent with several models of how foveal and peripheral phase encoding differ^{6,7,14}. For example, Rentschler and Treutwein⁷ proposed that two processes mediate grating discrimination. One process, operating only in the fovea, registers the relative positions of features. A second process, operating in the fovea and periphery, registers changes

in pattern contrast. Therefore, only phase shifts that alter contrast should be discriminable in the periphery. Braddick's model⁶ makes a similar prediction. The phase shifts we used did not alter contrast, so these models^{6,7} predict that none of the shifts should be discriminable in the periphery. Our results are clearly inconsistent with this prediction. Watt's model¹⁴ predicts that peripheral phase-encoding deficits will be reduced by increasing stimulus size. Again, our results are inconsistent with this prediction.

We examined whether the two-channel model can predict Rentschler and Treutwein's⁷ findings. They measured phase discrimination thresholds in the fovea and near periphery with compound gratings composed of a fundamental (F) and third harmonic ($3F$). The contrast of $3F$ was fixed at one-third that of F , but the phase of $3F$ was shifted around a mean phase of 0° , 90° , or 180° . Foveal thresholds in the 0° and 180° conditions were somewhat higher than those in the 90° condition. In the near periphery, however, shifts about 0° and 180° could not be discriminated. Only shifts about 90° were discriminable. The two-channel model predicts the peripheral results accurately, but it also predicts that phase shifts around 0° and 180° should not be discriminated in the fovea. We can account for the foveal results by altering the model slightly. Instead of using the responses exclusively of mechanisms centred precisely on the peaks, troughs, and zero-crossings of the waveform, the responses of slightly decentred mechanisms were used. Mechanisms centred on points up to 40 s on either side of a peak, trough, or zero-crossing enable foveal phase discriminations about 0° and 180° . With this new (and more biologically plausible) assumption, the two-channel model predicts Rentschler and Treutwein's findings quite accurately. Specifically, it predicts that foveal discrimination thresholds in the 0° and 180° conditions should be somewhat higher than in the 90° condition, and that thresholds in the 0° and 180° conditions, but not in the 90° condition, should rise substantially in the near periphery. The added assumption does not alter the predictions for our experiment, provided that the decentring is small.

In summary, our experiments are consistent with previous reports of phase-encoding anomalies in the periphery. They suggest, however, that the anomalies stem from sensitivity losses among odd-symmetric, but not even-symmetric, mechanisms. It is interesting to speculate that many of the recognition and identification deficiencies observed in peripheral vision may be caused in large part by changes in the sensitivity of one type of

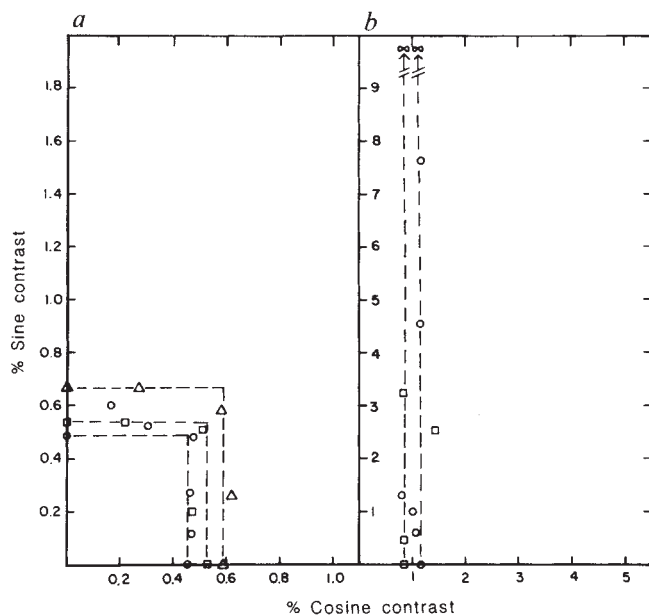


Fig. 4 The effect of spatial frequency on phase discrimination. *a*, Foveal phase discrimination thresholds: $\circ$, $F = 4$ c per deg, $C_F = 10\%$; $\square$, $F = 8$ c per deg, $C_F = 12.5\%$; $\triangle$, $F = 12$ c per deg, $C_F = 15\%$. Sensitivities to 0–180 and 90–270 shifts were approximately similar at all spatial frequencies. *b*, Discrimination thresholds at 40°: $\circ$, $F = 0.5$ c per deg, $C_F = 10\%$; $\square$, $F = 0.167$ c per deg, $C_F = 15\%$. Lowering spatial frequency does not alter the pattern of results. In particular, it was impossible to discriminate a 90–270° phase shift at both spatial frequencies. Note the change of scale on both axes between *a* and *b*.

mechanism, and a corresponding inability to pick up one class of phase information. This hypothesis is difficult to test currently because recognition and identification models are not sufficiently detailed to make quantitative predictions. However, even and odd mechanisms have been postulated in quantitative models of foveal grating detection and discrimination^{15–20}, motion discrimination^{21–23}, and the perception of Mach bands²⁴. If the mechanisms proposed in these models are the same as the even and odd mechanisms of our two-channel model, we predict that particular peripheral deficits will be observed in these visual processes. Such a prediction for the perception of Mach bands was confirmed recently²⁴. It remains to be seen if the other predictions are confirmed.

This research was funded by a research grant from NIH to M.S.B. We thank Jacob Nachmias for helpful comments on an earlier draft.

Received 18 December 1986; accepted 24 February 1987.

1. Rovamo, J., Virsu, V. & Nasanen, R. *Nature* **271**, 54–56 (1978).
2. Koenderink, J. J., Bouman, M. A., Bueno de Mesquita, A. E. & Slappendael, S. *J. opt. Soc. Am.* **68**, 854–860 (1978).
3. Merchant, J. *J. opt. Soc. Am.* **55**, 1291–1295 (1965).
4. Loomis, J. *Vision Res.* **18**, 335–338 (1978).
5. Julesz, B. *Nature* **290**, 91–97 (1981).
6. Braddick, O. *Documenta ophth. Proc. Series* **30**, 255–262 (1981).
7. Rentschler, I. & Treutwein, B. *Nature* **313**, 308–310 (1985).
8. Field, D. J. & Nachmias, J. *Vision Res.* **24**, 333–340 (1984).
9. Stromeyer, C. F. & Julesz, B. *J. opt. Soc. Am.* **62**, 1221–1232 (1972).
10. Nachmias, J. & Weber, A. *Vision Res.* **15**, 217–224 (1975).
11. Tolhurst, D. J. & Barfield, L. P. *Vision Res.* **18**, 951–958 (1978).
12. Levi, D. M., Klein, S. A. & Aitsebaomo, A. P. *Vision Res.* **25**, 963–977 (1985).
13. Livingston, M. S. & Hubel, D. H. *Nature* **315**, 285 (1985).
14. Watt, R. J. *Nature* **313**, 266–267 (1985).
15. Stromeyer, C. F. & Klein, S. *Vision Res.* **14**, 1409–1420 (1974).
16. Stromeyer, C. F., Lange, A. F. & Ganz, L. *Vision Res.* **13**, 2345–2360 (1973).
17. Shapley, R. M. & Tolhurst, D. J. *J. Physiol., Lond.* **229**, 165–183 (1973).
18. Tolhurst, D. J. *Vision Res.* **12**, 797–804 (1972).
19. Lawton, T. B. *Vision Res.* **24**, 139–148 (1984).
20. Lawton, T. B. *J. opt. Soc. Am. A* **2**, 1140–1152 (1985).
21. Van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am. A* **2**, 185–212 (1985).
22. Nakayama, K. & Silverman, G. H. *J. opt. Soc. Am. A* **2**, 267–274 (1985).
23. Watson, A. B. & Ahumada, A. J. *J. opt. Soc. Am. A* **2**, 322–342 (1985).
24. Morrone, M. C., Ross, J., Burr, D. C. & Owens, R. *Nature* **324**, 250–253 (1986).

Separation of the genetic loci for the H-Y antigen and for testis determination on human Y chromosome

Elizabeth Simpson*, Phillip Chandler*, Els Goulmy†, Christine M. Disteche‡, Malcolm A. Ferguson-Smith§ & David C. Page||

*Transplantation Biology Section, Clinical Research Centre, Harrow, Middlesex HA1 3UJ, UK

†Department of Immunohaematology and Blood Bank AZL, University Hospital Leiden, The Netherlands

‡Department of Pathology, University of Washington, Seattle, Washington 98195, USA

§Duncan Guthrie Institute of Medical Genetics, University of Glasgow, Glasgow, G3 8SJ, UK

||Whitehead Institute, Cambridge Massachusetts 02142, USA

The mammalian Y chromosome encodes a testis-determining factor (termed TDF in the human), a master regulator of sex differentiation. Embryos with a Y chromosome develop testes and become males whereas embryos lacking a Y chromosome develop ovaries and become females. Expression of H-Y, a minor histocompatibility antigen, may also be controlled by a gene on the Y chromosome, and it has been proposed that this antigen is the testis-determining factor¹. We have tested the postulated identity of H-Y and TDF in the human. H-Y typing with T cells was carried out on a series of sex-reversed humans (XX males and XY females), each shown by DNA hybridization to carry part but not all of the Y chromosome. This deletion analysis maps the gene for H-Y to the long arm or centromeric region of the human Y chromosome, far from the TDF locus, which maps to the distal short arm.

H-Y typing of sex-reversed mice has shed light on the relationship of this antigen to gonadal sex determination. XX *Sxr* (sex reversed) mice are both male and H-Y positive owing to the presence of a small portion of the Y chromosome^{2,3}, consistent with a testis-determining role for H-Y. However, a derivative of *Sxr* known as *Sxr'* produces XX *Sxr'* and XO *Sxr'* mice that are phenotypically male but H-Y negative^{4,5}, demonstrating that the antigen is not required for testis determination.

In humans, DNA hybridization studies using Y-specific probes of individuals with only parts of the normal Y chromosome have yielded an eight-interval deletion map of the Y chromosome. Such studies of XX males and XY females that carry part but not all of the Y chromosome, have demonstrated that interval 1, representing a portion of the distal short arm of the Y, contains TDF^{6,7}. Similar studies of another series of XX males, using a different set of Y-DNA probes, likewise show them to possess sequences both within and distal to Yp11.2 (refs 8 and 9). We reasoned that if there is a Y-chromosomal gene responsible for H-Y antigen expression then it should be possible to map that gene by H-Y typing individuals with well-characterized Y deletions.

H-Y-specific T-cell clones and lines have been derived *in vitro* from transfused human female aplastic anaemia patients¹⁰. These H-Y-specific T cells, whose response is HLA restricted, were used to H-Y type Epstein-Barr-virus-transformed B-cell lines from deleted-Y individuals. Because only HLA-A2 or B7-restricted H-Y-specific T-cell clones were available at the time of this study, we were able to H-Y type only those deleted-Y individuals who were HLA-A2 and/or -B7. Variants of HLA-A2 and -B7 which cannot be recognized by allogeneic or HLA-restricted T cells have been described¹¹. Accordingly, in each case, the HLA type was determined serologically and confirmed using alloreactive cytotoxic T cells.

Table 1 shows the results of HLA and H-Y typing of eight deleted-Y individuals (six 46,XX males and two 46,XY females). All six XX males were negative for H-Y. These include three

Table 1 HLA and H-Y typing of B-cell lines from XX males, XY females and normal controls

Experiment	Karyotype/sex	Individual	HLA serology*	Specific lysis (%)		H-Y phenotype
				α A-2	α H-Y/A-2	
1	XX♂	RH	A2	<u>18</u>	9	—
	XX♂	JT	A2	<u>24</u>	3	—
	XX♂	LGL 105	A2	<u>13</u>	4	—
	XY♂	Normal male	A2	<u>20</u>	<u>38</u>	+
	XX♀	Normal female	A2	<u>17</u>	8	—
2	XX♂	WB	A2	<u>37</u>	0	—
	XY♂	Normal male	A2	<u>25</u>	<u>17</u>	+
	XX♀	Normal female	A2	<u>17</u>	3	—
3	XX♂	WHT950	B7	<u>αB7</u>	<u>αH-Y/B7</u>	—
	XX♂	JM	B7	<u>76</u>	9	—
	XY♂	Normal male	B7	<u>62</u>	1	—
	XX♀	Normal female	B7	ND	<u>40</u>	+
	XX♀	Normal female	B7	<u>54</u>	0	—
4	XY♀	WHT 1003 (case 1)	B7	<u>55</u>	<u>70</u>	+
	XY♀	WHT 715 (case 2)	B7	<u>57</u>	<u>69</u>	+
	XY♂	Father of case 2	B7	<u>52</u>	<u>61</u>	+
	XX♀	Mother of case 2	B7	<u>36</u>	6	—

The H-Y-specific cytotoxic T cells were T-cell lines or clones derived from the peripheral blood of transfused female aplastic anaemia patients¹⁰. The alloreactive cytotoxic T cells were derived from the peripheral blood of healthy individuals¹¹. They were used in the cytotoxic assays 5–7 days after aliquots previously stored in liquid nitrogen had been thawed and restimulated *in vitro* with appropriate antigen bearing irradiated peripheral blood cells in the presence of interleukin-2 (IL-2) in RPMI medium containing penicillin, streptomycin, glutamine, HEPES and 10% fetal calf serum (FCS) or human serum. The target cells were Epstein-Barr virus (EBV)-transformed B-cell lines grown *in vitro*. Immediately before the assay, aliquots were labelled with ⁵¹Cr-labelled sodium chromate, washed and dispensed at 5×10^3 cells per well in round-bottomed microtitre wells. Triplicate wells were used at each A: T ratio^{3,4,14}. Per cent specific lysis given is that at an attacker-to-target ratio of 10:1, the figure being taken from a regression analysis of a three or four point titration curve. Underlined figures are positive values on titration curves whose r^2 value lay between 0.70 and 1.00 (refs 3, 4 and 14). ND, Not done.

* HLA serology of the B-cell lines from sex-reversed patients and the parents of one was performed either by Lorna Kennedy at the ICRF Lincoln's Inn Field, or by Donald Palmer in the Department of Immunology, Hammersmith Hospital. Normal male and female cells in experiments 1–3 were EBV-transformed B-cell lines from Professor A. Rickinson of Birmingham: the individuals from which they had been obtained were serologically identified as A2 or B7 by him.

XX males, LGL105, WHT950 and WB, who carry intervals 1, 2 and 3 of the Y chromosomal short arm (refs 6 and 12; D.C.P. unpublished results; see class 3 XX males in Fig. 1) and three XX males, RH, JT and JM, with similar portions of the Y short arm detected using a different set of Y-DNA probes^{8,9}. Whereas *TDF* is in interval 1 of the Y chromosome^{6,7}, these results indicate that the gene for H-Y lies outside intervals 1, 2 and 3.

This conclusion is reinforced by the finding that both XY females are positive for H-Y. XY female WHT1003 has a deletion of intervals 1, 2 and 3 on the Y short arm, whereas XY female WHT715 has a deletion of intervals 1, 2 and 4A (refs 7 and 13; see class 1 and 2 XY females in Fig. 1). Taken together, these results with XX males and XY females strongly suggest that the gene for H-Y maps somewhere in intervals 4B to 7 of the Y chromosome, that is, near the centromere or on the long arm (refs 6 and 7; Fig. 1).

The finding that XX *Sxr'* male mice were H-Y negative strongly suggested that H-Y is not the testis determinant. However, it might be that the *Sxr'* mutation had affected the antigenicity but not the testis-determining capability of H-Y. There is an earlier report of human XX males with negative H-Y type and an XY female with positive H-Y type with cytotoxic T cells¹⁴. However, the XX males had not been tested for the presence of Y-chromosomal DNA and the XY female had not been tested for deletions of Y DNA sequences. The results reported here are not subject to such ambiguities of interpretation. In the present study, H-Y typing was carried out on cells from eight sex-reversed individuals with well-characterized deletions of portions of the Y chromosome. The combination of negative and positive findings reported here demonstrates unequivocally that the gene for H-Y maps far from interval 1, which contains *TDF* and that H-Y antigen is not the testis-determining factor. H-Y antigen was originally defined by graft rejection¹⁵. The *in vitro* T-cell assays of the type used for H-Y

typing in this study have been shown to recognize the same histocompatibility antigen as that defined by graft rejection¹⁶. In contrast, serological identification of a male antigen assumed to be H-Y (and now called serologically detected male antigen, SDM) has been problematical. In particular, there is doubt concerning whether the SDM antigen is identical to the H-Y histocompatibility antigen^{5,17}. The implications of the present study, therefore, are limited to the H-Y antigen and do not extend to the SDM antigen.

We are currently examining additional deleted-Y individuals in an effort to map the H-Y locus more precisely. The possibility that the H-Y gene might be identical to or near a spermatogenesis factor on the long arm of the human Y chromosome¹⁸ is of particular interest, because in mice H-Y (or a closely linked gene) has been implicated in spermatogenesis¹⁹. Though the results reported here imply that H-Y has no function in gonadal sex determination, it may be involved in testis function at a later stage of development.

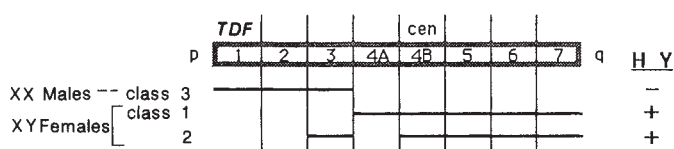


Fig. 1 Eight-interval deletion map of the human Y chromosome (based on refs 7 and 13). The results of H-Y typing of individuals (from Table 1) are summarized. The short arm, centromere and long arm of the Y chromosome are denoted by, respectively 'p', 'cen' and 'q'. *TDF* gene for testis determining factor. Horizontal lines, portions of the chromosome present in the designated individual.

We thank Els Blokland and Jos Pool for isolating H-Y-specific T-cell lines and clones and for assistance with the typing; Albert de la Chapelle, Joseph Hersh and Martin Crawford for providing cell lines and blood samples; and Lesley Snadden for help in setting up and maintaining EB transformed lymphoblastoid lines. This work was supported in part by grants from the MRC (MAFS), the NIH, the March of Dimes, the American Cancer Society, the Dutch Foundation for Medical Health Research (Medigon), which is subsidized by the Dutch Organisation for the Advancement of Pure Research (ZWO) and the J. A. Cohen Institute for Radiopathology and Radiation Protection (IRS).

Received 3 February; accepted 5 March 1987.

- Wachtel, S. S., Ohno, S., Koo, G. C. & Boyse, E. A. *Nature* **257**, 235-236 (1975).
- Bennett, D. et al. *Nature* **265**, 255-257 (1977).
- Simpson, E., Edwards, P., Wachtel, S. S., McLaren, A. & Chandler, P. *Immunogenetics* **13**, 355-358 (1981).
- McLaren, A., Simpson, E., Tomonari, K., Chandler, P. & Hogg, H. *Nature* **312**, 552-555 (1984).
- Simpson, E. *Cell* **44**, 813-814 (1986).
- Vergnaud, G. et al. *Am. J. hum. Genet.* **38**, 109-124 (1986).
- Page, D. C. *Cold Spring Harb. Symp. quant. Biol.* **51**, 229-235 (1986).
- Affara, N. A. et al. *Nucleic Acids Res.* **14**, 5353-5373 (1986).
- Affara, N. A. et al. *Nucleic Acids Res.* **14**, 5375-5387 (1986).
- Goulmy, E. *Progress in Allergy* **36**, 44-72 (1985).
- Horai, S., van der Poel, J. J. & Goulmy, E. *Immunogenetics* **16**, 135-142 (1982).
- Andersson, M., Page, D. C. & de la Chapelle, A. *Science* **233**, 786-788 (1986).
- Disteche, C. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7841-7844 (1986).
- Goulmy, E., van Leeuwen, A., Blokland, E., Sachs, E. S. & Geraedts, J. P. M. *Immunogenetics* **17**, 523-531 (1983).
- Eichwald, E. J. & Siltser, C. R. *Transplant. Bull.* **2**, 148 (1955).
- Simpson, E. et al. *Immunology* **57**, 345-349 (1986).
- Silvers, W. K., Gasser, D. L. & Eichler, E. M. *Cell* **28**, 439-440 (1982).
- Tiepolo, L. & Zuffardi, O. *Hum. Genet.* **34**, 119-124 (1976).
- Burgoyne, P. S., Levy, E. R. & McLaren, A. *Nature* **320**, 170-172 (1986).

Tumour prevention and rejection with recombinant vaccinia

R. Lathe*, M. P. Kieny†, P. Gerlinger*, P. Clertant‡, I. Guizani‡, F. Cuzin‡ & P. Chambon*

*CNRS-LGME and U184-INSERM, 11 Rue Humann, 67085 Strasbourg, France

†Transgène S.A., 11 Rue de Molsheim, 67000 Strasbourg, France

‡U273-INSERM, Parc Valrose, Nice, France

Tumour-specific antigens (TSA; ref. 1) have been exploited in the diagnosis² and imaging³ of human cancer and anti-TSA antibodies have therapeutic potential⁴. Vaccination with TSA⁵⁻⁸ or anti-idiotypic (TSA) antibodies^{9,10} has also been used to control tumour growth in model systems. An effective immune response nevertheless demands copresentation of antigen with host histocompatibility determinants¹¹. We therefore examined whether live vaccinia virus recombinants expressing TSA in cells of the vaccinated host might better elicit tumour immunity. Polyoma virus (PY) is tumorigenic in rodents; because killed PY-transformed cells can elicit tumour immunity, a PY-specific TSA has been postulated¹²⁻¹⁴. Tumorigenesis involves expression of three early PY proteins¹⁵⁻¹⁷, large-T (LT), middle-T (MT) and small-T (ST), but their role as TSAs is unclear. We therefore expressed the three T proteins in separate vaccinia recombinants. Rejection of PY tumours was observed in rats immunized with recombinants expressing either LT or MT. Further, tumour-bearing animals could be induced to reject their tumours by inoculation of recombinants.

Live vaccinia recombinants^{18,19} expressing foreign antigens have been used successfully in immunization experiments (see ref. 20 for review) and we have previously applied this approach to a rabies vaccine^{21,22}. Vaccinia seemed appropriate for the expression of tumour antigens because recombinants can prime a cell-mediated immune response²²⁻²⁵ important in tumour cell elimination^{26,27}.

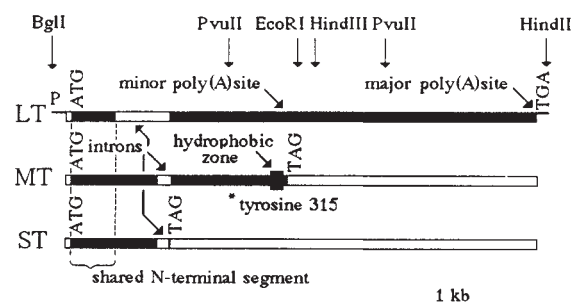


Fig. 1 Polyoma early region and cloning of coding sequences into vaccinia virus. Schematic representation of the PY early region extending from the unique *Bgl*I site to the *Hind*II site (adapted from ref. 17). The promoter for early transcription (P), the translation initiation (ATG) and termination (TGA, TAG) signals and the polyadenylation sites are shown. Solid bar, coding regions; open bar, introns (excised from mature messenger) and untranslated regions. The zone encoding the carboxy-terminal hydrophobic, potentially transmembrane, region of MT is indicated. Tyr 315 is the major site at which MT is phosphorylated in PY-transformed cells.

Methods. For construction of recombinants, intron-free coding sequences for the three T proteins were excised from plasmids pPY-LT1, pPY-MT1 and pPY-ST1 (ref. 29) with *Bgl*I at one end and *Hind*II, *Eco*RI and *Pvu*II respectively at the other. Using a single-stranded oligonucleotide adaptor converting the *Bgl*I extremity to compatibility with ends generated by *Bam*HI cleavage, the three coding sequences were separately introduced between the *Bam*HI-*Sma*I, *Bam*HI-*Eco*RI and *Bam*HI-*Sma*I sites respectively of plasmid pTg186-poly⁴⁷, which is pTg1H-TK-7.5K (ref. 21) in which a polylinker has been introduced downstream of the vaccinia 7.5K gene promoter in the body of the viral thymidine kinase (TK) gene. Procedures for the propagation and purification of vaccinia virus and for transfer of coding sequences and accompanying promoter to the vaccinia genome were as described²¹.

In the PY genome the genes for LT, MT and ST overlap and messenger RNAs are generated by differential splicing (Fig. 1). DNA sequences separately encoding the three T proteins have been constructed^{28,29} and we introduced these separately downstream of the promoter from the vaccinia 7.5K gene. Recombination *in vivo* was used for transfer to the vaccinia genome, generating strains labelled VVpyLT, VVpyMT, and VVpyST according to the T protein concerned.

To confirm the authenticity of the recombinant T species, proteins synthesized on infection of tissue-culture cells with the recombinants were examined by Western blotting for reaction with anti-T serum. Protein bands recognized by anti-T serum had electrophoretic mobilities expected for authentic PY proteins (Fig. 2). In all cases the recombinant T proteins possess biochemical characteristics associated with the cognate PY polypeptides (P.C., I.G., M.P.K. and R.L., in preparation).

To demonstrate appropriate sub-cellular distribution of the three T proteins, infected cells were examined for reaction with fluorescently labelled anti-T serum. Despite alterations in cellular morphology on vaccinia infection, the LT protein was, as expected^{30,16}, observed only in the nucleus (dark nucleoli, Fig. 3a) whereas the ST species was cytoplasmic (Fig. 3c). Fluorescence in cells infected with VVpyMT revealed cytoplasmic aggregates (Fig. 3b) as well as perinuclear staining that may suggest an association with intracellular membranes^{31,32}. Cell-surface MT^{14,33} was not observed.

Rats inoculated subcutaneously with PY-transformed syngeneic rat cells rapidly develop tumours. We sought to determine whether prior administration of the vaccinia/polyoma recombinants blocked tumour development. Animals were inoculated with recombinant virus, boosted, and seeded with syngeneic PY-transformed cells (Table 1 legend). In view of a possible cumulative action of the different T proteins (see, for

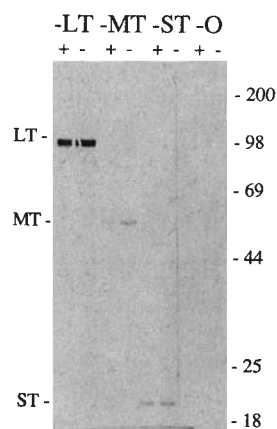
Table 1 Tumour rejection in animals inoculated with vaccinia recombinants

Virus administered	Route	Animals showing tumour rejection	Total number of animals	Rejection frequency
Vaccination before tumour inoculation				
No virus	s.c.	0	6	0
VVpyST	s.c.	0	7	0
VVpyMT	s.c.	6	10	0.6
VVpyLT	s.c.	6	10	0.6
VVpyST, MT, ST	s.c.	2*	4	0.5
No virus	i.d.	0	8	0
VV-c	i.d.	0	4	0
VVpyST	i.d.	0	6	0
VVpyMT	i.d.	14	15	0.9
VVpyLT	i.d.	5	8	0.6
Tumour development before vaccination				
No virus	i.d.	0	4	0
VVpyST	s.c.	0	4	0
VVpyMT	s.c.	0	6	0
VVpyLT	s.c.	2	10	0.2
No virus	i.d.	0	4	0
VVpyST	i.d.	0	4	0
VVpyMT	i.d.	4	8	0.5
VVpyLT	i.d.	0	8	0

* One animal developed no confirmed tumour throughout the experiment. In a second animal, a small tumour (diameter ~3 mm) appeared only at 32 days and persisted until the animal was killed (45 days). Other vaccinated animals rejecting their tumours have subsequently remained tumour-free (up to 12 months).

Methods. Groups for four-week old female IOPS Fischer rats (IFFA-Credo) were inoculated (subcutaneously, s.c. with 10^7 plaque-forming units (PFU) of purified recombinant virus in a volume of 100 μ l and boosted with a similar dose at 12 days. For intradermal (i.d.) vaccination purified virus suspension (20 μ l, 10^7 PFU) was applied to abrasions at the base of the tail. The control virus, VV-c, was a vaccinia recombinant bearing an unrelated coding sequence (not described). Challenge at 16 days was with 2×10^4 polyoma (complete genome)-transformed Fischer rat 3T3 cells (PYT-21, ref. 29) in a volume of 100 μ l (the presence of all three PY T-proteins in PYT-21 cells has been confirmed; M. Rassoulzadegan, personal communication). For post-challenge vaccination animals were seeded with tumour cells as above and inoculated 11 and 15 days later.

Fig. 2 Vaccinia-encoded PY T proteins analysed by Western blotting. Mouse 3T6 cells were infected with recombinant viruses (multiplicity = 2), in the presence (+) or absence (-) of actinomycin C1 ($1 \mu\text{g ml}^{-1}$, 2 h after infection). Proteins were extracted after 12 h, separated by SDS-polyacrylamide gel electrophoresis and T proteins were revealed with anti-T serum and a peroxidase-labelled second antibody. Infecting viruses were VVpyLT, VVpyMT, VVpyST and VV-O (a spontaneous TK-negative non-recombinant vaccinia virus) as indicated by the suffix above each lane. Size standards (right) are relative molecular mass in thousands (M_r in K). The positions of the LT, MT and ST species are indicated with M_r : 105K; MT, 56K; ST, 22K. The inability of actinomycin C1 to block expression confirms the viral recombinant origin of the T proteins.



example, ref. 29) further animals were vaccinated with all three recombinants. All unvaccinated animals developed localized tumours and no spontaneous regression was observed during the experiments. In contrast, 50–60% of animals vaccinated (subcutaneously) with either VVpyLT or with VVpyMT developed small tumours (5–10 mm in diameter) which regressed and were eliminated (Table 1). Simultaneous administration of all three recombinants yielded no improvement in regression frequency. The route of vaccination proved important, and intradermal rather than subcutaneous inoculation increased the rejection rate (90% for the VVpyMT vaccine; Table 1). Although the polyoma T-proteins are known to participate in oncogenic transformation²⁹ the recombinant viruses were not themselves oncogenic (Table 1 legend).

We were concerned that vaccinia virus might retard tumour growth in a non-specific manner³⁴, or that vaccinia infection of

tumour cells might interfere with tumour development³⁵. However, animals receiving VVpyST or a vaccinia recombinant lacking polyoma sequences failed to reject their tumours (Table 1). Similarly, syngeneic tumour cells lacking PY sequences (FR3T3-ras) were not rejected after vaccination with VVpyLT or VVpyMT (data not shown).

To determine the role of antibodies against the T proteins in tumour rejection we examined sera from immunized animals for reaction with PY-transformed cells. Surprisingly, vaccinated animals uniformly presented antibody reacting with immobilized syngeneic PY-transformed rat cells. Sera from vaccinated animals also reacted strongly with the established syngeneic rat cell line FR3T3 (not shown), showing that inoculation with live vaccinia virus provokes a reaction against host components. Immunoaffinity studies (data not shown) revealed only weak reaction with T proteins and no reproducible differences in the

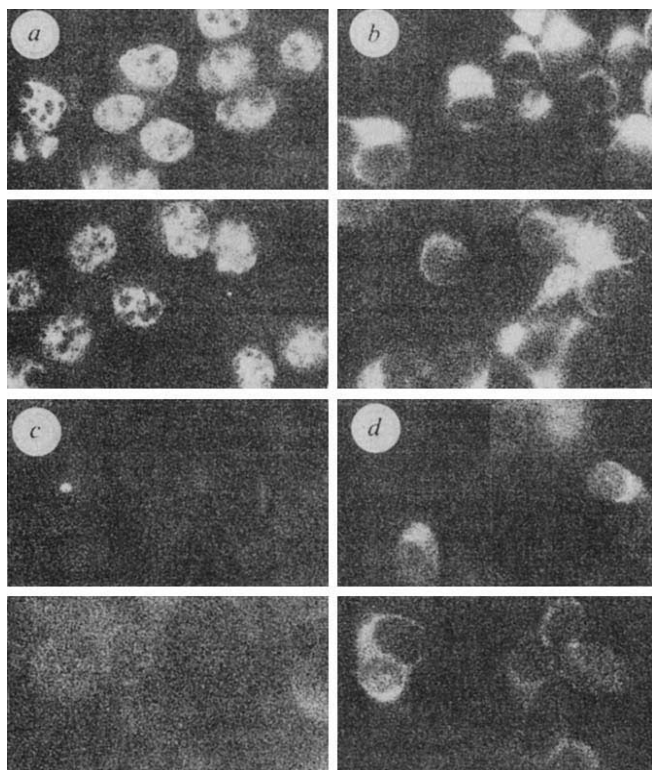


Fig. 3 Immunofluorescence of cells infected with vaccinia-T recombinant viruses. Rodent cells infected with VVpyLT (a), VVpyMT (b), VVpyST (d), or uninfected (c) were probed with anti-T serum and a fluorescent second antibody.

Methods. Semi-confluent monolayers of BHK-21 cells were infected with purified recombinant virus (0.1 PFU per cell) and incubated overnight. Cells were fixed with cold 80% acetone, dried and T polypeptides detected by the sequential application of rat anti-T immune serum derived from a tumour-bearing animal (RAF no. 4 (ref. 48) 1/50 in phosphate-buffered saline (PBS) with 1% bovine serum albumin (BSA); 20 min, 37 °C) and fluorescein isothiocyanate (FITC)-labelled goat anti-rat immunoglobulin (Miles; 1/100 in PBS + 1% BSA; 20 min, 37 °C).

antigen-binding profiles were observed. Although such tests may not reveal MHC-restricted antibody³⁶ we surmise that circulating anti-T antibody may not be responsible for tumour regression.

We also examined whether animals already bearing tumours could be induced to reject their tumour cells by vaccination with the appropriate recombinant. Rats were inoculated with PY-transformed cells and allowed to develop tumours. When all animals presented tumours (2–4 mm diameter) they were inoculated twice with the vaccinia recombinants. In all cases tumour growth continued to a diameter exceeding 15 mm. However, tumour growth halted in 4 of 8 animals receiving VVpyMT (intradermal) and regression proceeded to complete

tumour elimination (Table 1). Unvaccinated animals failed to reject their tumours (not shown).

Although vaccinia recombinants or killed cells³⁷ expressing LT or MT coding sequences can elicit tumour immunity, the precise identity of PY-specific TSA remains unknown. TSAs are generally assumed to be cell-surface molecules, but none of the PY early proteins possesses evident signals for cell-surface presentation, although MT does carry a carboxy-terminal hydrophobic (and thus potentially transmembrane) domain (Fig. 1). Because the coding sequences for MT and LT overlap, our results do not exclude the possibility that TSA may be a minor protein species¹⁶ produced from both VVpyLT and VVpyMT. However, it has been proposed that the related LT protein from simian virus 40 (SV40)^{16,17} reaches the cell surface through an alternative nuclear exocytotic pathway³⁸ and it is now becoming clear that 'internal' proteins can be processed for presentation to the immune system^{39–41}.

We have demonstrated, in a model system, that administration of live recombinant vaccinia virus bearing tumour-specific antigen can efficiently prevent the development of transplanted tumour cells. Further, a significant number of tumour-bearing animals can be induced to reject their tumours by vaccination. Appropriate live viral recombinants thus have considerable promise in the control of human tumours presenting novel antigens^{8,42–46}.

We thank Drs J. P. Lecocq, A. Kirn, G. Henri, M. Lemeur, T. Dalianis and H. Koprowski for discussions and D. Schmitt, K. Dott and F. Mackay-Landau for technical assistance.

Received 12 January; accepted 4 March 1987.

- Hellström, K. E., Hellström, I. & Brown, J. P. *Springer Semin. Immunopath.* **5**, 127–146 (1982).
- Herlyn, M., Blaszczyk, M. & Koprowski, H. *Contr. Oncol.* **19**, 160–170 (1984).
- Begent, R. H. J. *Biochem. biophys. Acta* **780**, 151–166 (1985).
- Miller, R. A. *et al. Hybridomas in Cancer Diagnosis and Treatment* (eds Mitchell, M. S. & Oettgen, H. F.) 133–140 (Raven, New York, 1982).
- Gross, L. *Cancer Res.* **3**, 326–333 (1943).
- Foley, E. J. *Cancer Res.* **13**, 835–837 (1953).
- Prehn, R. T. & Main, J. M. *J. natn. Cancer Inst.* **18**, 769–778 (1957).
- Tevethia, S. S., Flyer, D. C. & Tijan, R. *Virology* **107**, 13–23 (1980).
- Lee, V. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6286–6290 (1985).
- Herlyn, D., Ross, A. H. & Koprowski, H. *Science* **232**, 100–102 (1986).
- Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
- Sjogren, H. O., Hellström, I. & Klein, I. *Cancer Res.* **21**, 329–337 (1961).
- Habel, K. *Proc. Soc. exp. Biol. Med.* **106**, 722–725 (1961).
- Ito, Y., Brocklehurst, J. R. & Dulbecco, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4666–4670 (1977).
- Basilico, C. *Pharmac. Ther.* **26**, 235–272 (1984).
- Griffin, B. E. & Dilworth, S. M. *Adv. Cancer Res.* **39**, 183–268 (1983).
- Toozie, J. *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1981).
- Panicali, D. & Paoletti, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4927–4931 (1982).
- Mackett, M., Smith, G. L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7415–7419 (1982).
- Smith, G. L., Mackett, M. & Moss, B. *Biotechnol. Genet. Engng. Rev.* **2**, 383–407 (1984).
- Kieny, M. P. *et al. Nature* **312**, 163–166 (1984).
- Wiktor, T. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7194–7198 (1984).
- Wiktor, T. J., Kieny, M. P. & Lathe, R. *Appl. Virol. Res.* (in the press).
- Bennink, J. R., Yewdell, J. W., Smith, G. L., Moller, C. & Moss, B. *Nature* **311**, 578–579 (1984).

- McMichael, A. J., Michie, C. A., Gotch, F. M., Smith, G. L. & Moss, B. *J. gen. Virol.* **67**, 719–726 (1986).
- Hellström, K. E. & Hellström, I. *Adv. Cancer Res.* **12**, 167–223 (1969).
- Heberman, R. B. *Adv. Cancer Res.* **19**, 207–263 (1974).
- Treisman, R., Novak, U., Favalaro, J. & Kamen, R. *Nature* **292**, 595–600 (1981).
- Rassoulzadegan, M. *et al. Nature* **300**, 713–718 (1982).
- Takemoto, K. K., Malmgren, R. A. & Habel, K. *Virology* **28**, 485–488 (1966).
- Zhu, Z. *et al. J. Virol.* **51**, 170–180 (1984).
- Dilworth, S. M. *et al. EMBO J.* **5**, 491–499 (1986).
- Schaffhausen, B. S., Dorai, H., Arakere, G. & Benjamin, T. L. *Molec. cell. Biol.* **2**, 1187–1198 (1982).
- Old, L. J., Clarke, D. A. & Benacerraf, B. *Nature* **184**, 291–292 (1959).
- Koprowski, H. & Norton, T. W. *Cancer* **3**, 874–885 (1950).
- Froscher, B. G. & Klinman, N. R. *J. exp. Med.* **164**, 196–210 (1986).
- Dalianis, T., Ramqvist, T. & Klein, G. *Int. J. Cancer* **34**, 403–406 (1984).
- Sharma, S., Rodgers, L., Brandsma, J., Gething, M. J. & Sambrook, J. *EMBO J.* **4**, 1479–1489 (1985).
- Bennink, J. R., Yewdell, J. W. & Gerhard, W. *Nature* **296**, 75–76 (1982).
- Townsend, A. R. M., McMichael, A. J., Carter, N. P., Haddleston, J. A. & Brownlee, G. G. *Cell* **39**, 13–25 (1984).
- Yewdell, J. W., Bennink, J. R., Smith, G. L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1785–1789 (1985).
- Koprowski, H., Herlyn, M., Steplewski, Z. & Sears, H. F. *Science* **212**, 53–56 (1981).
- Real, F. X. *et al. J. exp. Med.* **160**, 1219–1233 (1984).
- Ueda, R. *et al. J. exp. Med.* **150**, 564–572 (1979).
- Drebin, J. A., Stern, D. F., Link, V. C., Weinberg, R. A. & Greene, M. I. *Nature* **312**, 545–548 (1984).
- Peto, R. & Zur Hausen, H. (eds) *Banbury Report 21, Viral Etiology of Cervical Cancer* (Cold Spring Harbor Laboratory, New York, 1986).
- Kieny, M. P. *et al. Bio/Technology* **4**, 790–795 (1986).
- Clertant, P., Gaudray, P., May, E. & Cuzin, F. *J. biol. Chem.* **259**, 15196–15293 (1984).

Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2

Frances Gotch*, Jonathan Rothbard†, Kevin Howland†, Alain Townsend* & Andrew McMichael*

* Nuffield Department of Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

† Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC1, UK

Both human and murine cytotoxic T cells (CTL) elicited in response to infection with influenza A viruses have been shown to be specific for internal viral proteins, such as the matrix and nucleoprotein¹⁻³. Individual CTL epitopes have been identified in the nucleoprotein by successfully substituting short synthetic peptides for the intact virus in the preparation of target cells in cytotoxicity assays⁴. The defined peptide epitopes have each been recognized by CTL in association with individual class I major histocompatibility complex (MHC) proteins, H-2D^b (ref. 4), H-2K^k (ref. 5), H-2K^d (Taylor, P. *et al.*, unpublished data) and HLA-B37 (refs 4, 6). A logical strategy to investigate the molecular details of the interaction between antigen and MHC class I proteins would be to define an epitope recognized by the MHC class I molecule HLA-A2. This is because the amino-acid sequence is known, several variants of A2 have been characterized and the protein has been purified and crystallized⁷⁻¹¹. Here we describe a peptide derived from the influenza matrix protein that is recognized by human CTL in association with the HLA-A2 molecule.

Previous work in our laboratory using recombinant vaccinia virus has demonstrated that HLA-A2-restricted CTL are specific for an epitope in the matrix protein rather than nucleoprotein of influenza virus³. Based on an empirical prediction algorithm¹², eight peptides corresponding to linear regions of the matrix protein sequence¹³ were synthesized and tested for their ability to be recognized by A2-restricted influenza matrix protein specific CTLs (Fig. 1). Each peptide contained a linear sequence composed of a glycine or a charged residue followed by either two or three hydrophobic residues, and ending with a polar amino acid or glycine. These patterns of either four or five amino acids exist in the majority of defined T-cell epitopes¹². There are sixteen such patterns in the matrix protein and the peptides

Table 1 Recognition of matrix peptide 55-73 by CTL from different donors

Donor	HLA type*	E:T	% Specific lysis†		
			A/X31	55-73	0
I.H.	A2, 3; B7, 62; C3, 7; DR2, 4	18:1	31	21	5
J.M.	2, ; B51, 62; C3; DR4	40:1	29	22	2
G.P.	A2, 24; B7, 44; C5, 7; DR7, 11	50:1	37	26	3
A.M.	A2 ⁺ , 32; B49, 60; C3, 7; DR4	50:1	75	8	8
D.C.	A1, 28; B8, 18; C7; DR3, 4	6:1	28	10	10
N.C.	A1, 24; B7, 44; C7; DR2, 7	60:1	42	8	5
L.G.	A1, 29; B37, 44; C6	25:1	25‡	2	3
N.R.	A1, 3; B8, 60; C3, 7; DR2, 3	40:1	15	0	2
C.B.	A1, 30; B7, 44; C7, 5; DR2, 4	40:1	28	10	13

* HLA types kindly determined by C. Taylor and A. Ting, Nuffield Department of Surgery, Oxford; the DR type of L.G. has not been determined.

† Tested on autologous lymphoblastoid cells.

‡ Tested on autologous target cells treated with NP peptide 335-349. A2⁺, A2 variant (A2.2) which is more basic on isoelectric focusing. E:T, effector to target ratio.

a

MSLLTEVETVLSIIPSGPLKAEIAQRLEDVFAKNTDLEVLMEWLKTRP
17-31 34-53

ILSPLTKGILGFVFTLTVPSEGLQRRRFVQNALNGNDPNNMDKAVKLY
55-73 72-88 89-107

RKLKREITFHGAKEISLSYAGALASCMGLIYNRMGAIVTEVAFGLVCAT
129-145

CEQIADSQRRSHRQMTVTNPLIRHENRMVLASTTAKAMEQMGSSSEQAA
182-195

EAMEVASQARQMVQAMIRIGTHPSSSAGLKNLDLENLQAYQKRMGVQMRPK
226-241

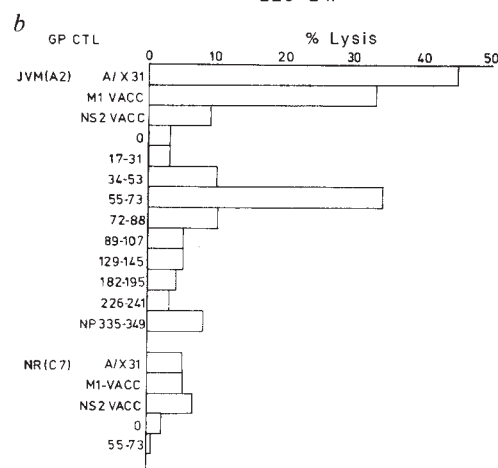


Fig. 1 Influenza A virus matrix protein and cytotoxic T lymphocyte recognition of peptides. *a*, Amino-acid sequence of the matrix protein¹³ using the single letter code for amino acids. The eight synthetic peptides made with an Applied Biosystems synthesizer are indicated. *b*, Lysis by influenza A virus specific CTL from donor G.P. of HLA-A2 matched (J.V.M.) or HLA-C7 matched (N.R.) lymphoblastoid cells that were infected with influenza A/X31 virus (A/X31), influenza matrix-1-expressing recombinant vaccinia virus (M1-VACC), influenza non-structural protein-2 expressing recombinant vaccinia virus (NS2-VACC), uninfected (0), or were treated with the synthetic peptides shown. The peptide NP 335-349 is derived from the nucleoprotein sequence and has been shown to be restricted by HLA-B37^{4,6}. Recombinant vaccinia viruses were gifts from Dr G. Smith and Dr B. Moss³.

Methods. Influenza A virus specific CTL were generated from peripheral blood lymphocytes of donor G.P. as described previously^{3,6}. After culture for seven days with A/X31 virus, they were tested on effector cells by mixing with 10^4 ⁵¹Cr-labelled target cells at a ratio of 40:1. Peptides were added to a final concentration of $20 \mu\text{g ml}^{-1}$ ($\sim 40 \mu\text{M}$). Supernatants were harvested after five h and specific lysis calculated from the formula: $(E - M/D - M) \times 100$. E, experimental ⁵¹Cr release. M, release in presence of culture medium, D, release by 5% Triton X-100.

described in this report include eleven. The remaining five patterns were included in two additional peptides which were assayed less extensively, but with negative results (data not shown).

Each peptide was assayed by incubation with either ⁵¹Cr-labelled autologous B lymphoblasts or lymphoblasts matched with the effector cells only through HLA-A2. Of the peptides tested, only the region corresponding to residues 55-73 was recognized by influenza A virus-specific CTL from donor GP (HLA type A2, 24; B7, 44; C7), (Fig. 1). Influenza-specific CTL

Fig. 2 Recognition of matrix peptide 55-73 is restricted by HLA-A2. Influenza A virus specific CTL from donor IH were tested on lymphoblastoid cell lines infected with influenza A/X31 virus (●—●), uninfected cells in the presence of peptide 55-73 (○—○), or uninfected cells in the absence of peptide (■—■). HLA antigens shared with the effector cells are shown in parentheses. HLA-A2⁺ is a variant of HLA-A2 that has a more basic isoelectric point¹⁴.

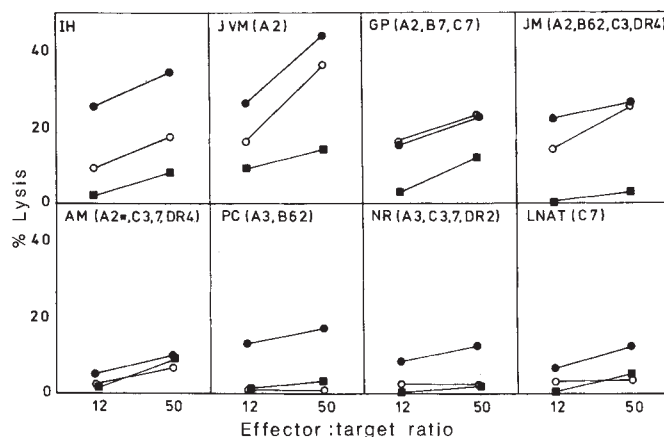
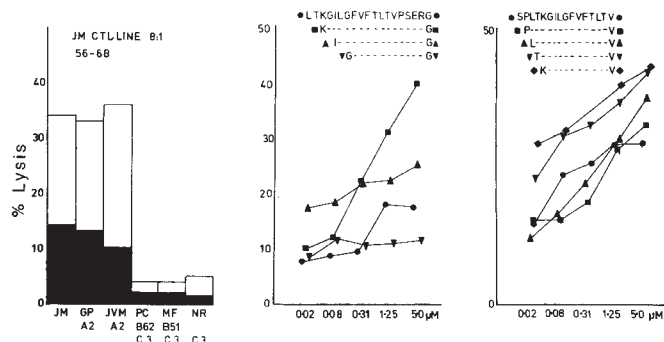


Fig. 3 Reactivity of J.M. CTL line restimulated with peptide 56-68. In the left-hand panel J.M. CTL were tested on lymphoblastoid cells at effector:target ratios of 8:1. The shared HLA antigens for each target are indicated. Each column represents lysis of each target in the presence of peptide (20 $\mu\text{g ml}^{-1}$); the solid column represents lysis in the absence of added peptide. The centre and right-hand panels demonstrate reactivity of J.M. CTL line at a killer:target ratio of 5:1 on autologous cells tested with the peptides shown at concentrations of 0.02–50 μM . Peptides, at these concentrations, were present throughout the cytolysis assay.



from two other A2-positive donors recognize this peptide, whereas CTL from six A2-negative donors did not (Table 1). One of these six was typed as an A2 variant (A2.2)¹⁴, and one as A28 which is also closely related to A2 in amino-acid sequence⁷. These results show that the dominant influence on whether an individual makes a CTL response to this particular influenza epitope is the presence or absence of HLA-A2. A similar conclusion was drawn from our observation that the NP peptide 335-349 was recognized in association with HLA-B37 by influenza specific CTL from five out of five unrelated donors with HLA-B37 (ref. 6).

The association between CTL recognition of peptide 55-73 and HLA-A2 was apparent when influenza-specific CTL from each A2-positive donor was tested, in the presence of the peptide, on target cells matched by single HLA antigens. In each case (see Fig. 2 for example), only those that expressed HLA-A2 were lysed.

A CTL line was prepared from influenza-virus-stimulated cells of the HLA-A2 positive donor J.M. by restimulating with irradiated autologous lymphoblastoid cells which had been pretreated with matrix peptide 56-68 and interleukin-2. This peptide was used in preference to 55-73 because preliminary experiments had indicated that it was recognized more effectively by J.M. CTL. J.M. CTL line cells lysed only target cells that had been pretreated with peptide 56-68 (or 55-73) and which shared HLA-A2 (Fig. 3). The T-cell line was also tested on two sets of truncated peptides (Fig. 3). For the set ending at valine 68, peptides starting at threonine 56 and lysine 57 were recognized most effectively. For the set ending at glycine 73, the optimal peptide started at lysine 57. Recognition was impaired by removal of glycine 58 although background lysis was raised. Removal of isoleucine 59 and leucine 60 prevented recognition by this CTL line. These three amino acids form part of the predictive pattern glycine, isoleucine, leucine, glycine. These results suggest that for HLA-A2-restricted CTL, the amino acids forming this pattern could be required in the peptide for CTL recognition to occur.

These and previous⁶ results indicate that the HLA class I molecules have a dominant role in determining which epitopes of influenza virus are recognized by CTL. HLA-A2 could exert this effect either by influencing the T-cell receptor repertoire during T-cell development or by selective binding of the matrix protein peptide produced during infection with influenza virus. These possibilities are not mutually exclusive. Recent work has demonstrated that MHC class II molecules bind antigenic peptides with low affinities and that the specificities of these reactions correlate with immune response gene phenotype¹⁵⁻¹⁸. By analogy, our results suggest that class I molecules and peptides may be found to interact similarly. The HLA-A2-restricted matrix peptide 56-68 should be especially valuable for testing this hypothesis. The putative interaction with purified HLA-A2 molecules can now be studied both in solution and the crystal.

We thank Dr J. Skehel for helpful discussions. This work was supported by the Medical Research Council and Imperial Cancer Research Fund.

Received 16 February; accepted 20 March 1986.

1. Townsend, A. R. M. & Skehel, J. J. *J. exp. Med.* **160**, 552-563 (1984).
2. Yewdell, J. W., Bennink, J. R., Smith, G. L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1785-1789 (1985).
3. Gotch, F. M., McMichael, A. J., Smith, G. L. & Moss, B. *J. exp. Med.* **165**, 408-416 (1987).
4. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
5. Bastin, J. M., Rothbard, J., Davey, J. & Townsend, A. R. M. *J. exp. Med.* (in the press).
6. McMichael, A. J., Gotch, F. M. & Rothbard, J. *J. exp. Med.* **164**, 1397-1406 (1986).
7. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3813-3817 (1982).
8. Krangel, M. S., Taketani, S., Biddison, W. E., Strong, D. M. & Strominger, J. L. *Biochemistry* **21**, 6313-6321 (1982).
9. Koller, B. H. & Orr, H. T. *J. Immun.* **134**, 2727-2733 (1985).
10. Taketani, S. M., Krangel, M. S., Pious, D. & Strominger, J. L. *J. Immun.* **131**, 2935-2938 (1983).
11. Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *J. molec. Biol.* **186**, 205-210 (1985).
12. Rothbard, J. B. & Taylor, W. *Nature* (submitted).
13. Winter, G. & Fields, S. *Nucleic Acids Res.* **8**, 1965-1974 (1980).
14. Gotch, F. M. *et al. Immunogenetics* **21**, 11-23 (1985).
15. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
16. Watts, T. H., Gaub, H. E. & McConnell, H. M. *Nature* **320**, 179-181 (1986).
17. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
18. Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071-1077 (1986).

Integral membrane glycoprotein properties of the prohormone pro-transforming growth factor- α

Joaquin Teixidó, Reid Gilmore, David C. Lee†* & Joan Massagué†

Department of Biochemistry, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

* Lineberger Cancer Research Center, University of North Carolina, Chapel Hill, North Carolina 27514, USA

Transforming growth factor- α (TGF- α) is a mitogenically active polypeptide hormone produced by acutely proliferating neoplastic and embryonic tissues¹⁻⁴. Molecular cloning of TGF- α cDNA from human and rat has indicated that this factor is synthesized as part of a larger precursor, proTGF- α ^{5,6}. An intriguing feature of proTGF- α is that it contains a sequence of 24 hydrophobic amino acids following the bioactive TGF- α sequence. Similar hydrophobic sequences present in transmembrane proteins are cotranslationally anchored into the phospholipid bilayer of the rough endoplasmic reticulum^{7,8}, suggesting the possibility that proTGF- α might be an integral membrane protein. Other secretory proteins including all known prohormones lack membrane anchoring sequences and are completely translocated across the endoplasmic reticulum membrane^{7,8}. To address the question of whether proTGF- α is as an integral membrane protein we have translated rat proTGF- α mRNA transcripts in the presence of rough endoplasmic reticulum membrane vesicles. The results indicate that proTGF- α is synthesized as an integral membrane glycoprotein. The 50-amino-acid TGF- α sequence in proTGF- α is exposed to the fluid extracellular phase where it can be released after cleavage by the appropriate enzyme.

Rat proTGF- α messenger RNA was obtained by transcription of a rat proTGF- α complementary DNA⁶ inserted in the pGEM-4 vector. ProTGF- α mRNA was translated *in vitro* in a wheat germ translation system containing ³⁵S-cysteine, in the absence or presence of dog pancreas rough microsomal membrane vesicles. The resulting translation products were immunoprecipitated with either antibodies to TGF- α raised against a synthetic peptide that corresponds to the C-terminal portion of mature TGF- α (residues 71 to 88 in proTGF- α), or antibodies to proTGF- α raised against synthetic peptides that correspond to carboxy-terminal sequences of proTGF- α ⁹. In this translation system, proTGF- α mRNA directed the synthesis of a product of relative molecular mass 17,000 (M_r , 17K) that was recognized specifically by both antisera (Fig. 1A, lanes a, c). The M_r of this translation product was that expected for the 159-amino-acid rat proTGF- α . When microsomes were present during translation of proTGF- α mRNA, a 20K translation product was obtained in addition to the 17K product (Fig. 1A, lanes e, g). If the microsomes were recovered by centrifugation at the end of the translation reaction, both the 17K and the 20K products could be immunoprecipitated from the detergent-solubilized membrane pellet. The supernatant from this centrifugation contained only the 17K product (not shown, but see Fig. 3, lane a). Neither the 17K nor the 20K translation products were observed when proTGF- α mRNA was not included in the translation reaction (Fig. 1A, lanes i, j). In addition, immunoprecipitation in the presence of the synthetic peptides used to raise the antibodies demonstrated that both the 17K and 20K polypeptides were authentic proTGF- α (Fig. 1A, lanes b, d, f, h).

During translation-coupled translocation, a variety of enzymatic activities in the lumen of the endoplasmic reticulum can act on the nascent polypeptide. These activities include signal peptidase which removes the amino-terminal signal sequence

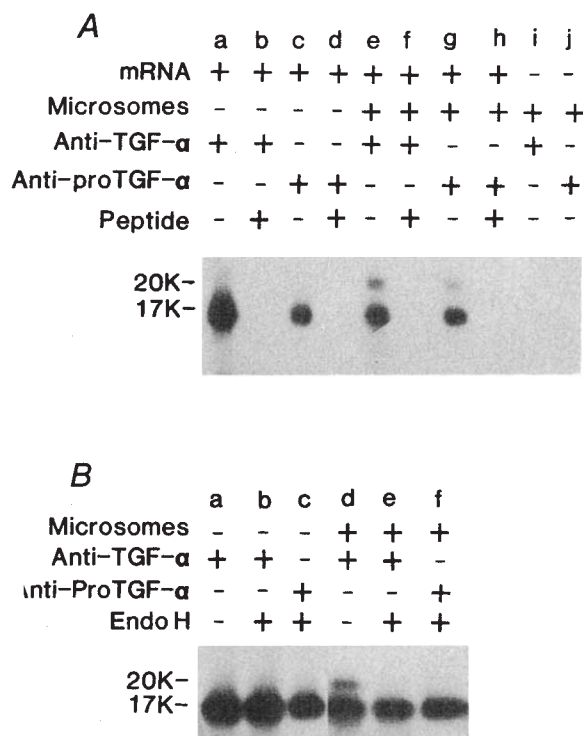


Fig. 1 Translation of rat proTGF- α mRNA in the presence or absence of rough microsomal membranes. **A**, Rat proTGF- α mRNA was translated in a wheat germ system alone (a-d) or with canine pancreas rough microsomal membranes (e-h). Translation products immunoprecipitated with anti-TGF- α or anti-proTGF- α antibodies. An excess of the corresponding synthetic peptide used to raise these antibodies was present during immunoprecipitation of some of the samples. Immunoprecipitates analysed by SDS-PAGE (polyacrylamide gel electrophoresis) and fluorography of the resulting gels. Negative controls (i, j) included 'translations' in the absence of proTGF- α mRNA. **B**, Some immunoprecipitates obtained after translation of proTGF- α , treated with endoglycosidase H before SDS-PAGE and fluorography.

Methods. The plasmid pK1 which contains the *Sma-Sma* fragment of rat proTGF- α cDNA⁶ with the 5' *Sma* site changed to a *Bam*HI site, was digested with *Bam*HI and *Ava*I (New England Biolabs). The resulting 717-base pair (bp) fragment which encompasses the full proTGF- α coding sequence was inserted in the vector pGEM-4 (Promega Biotec) to obtain the plasmid pGEM α . T7 RNA polymerase transcription of pGEM α linearized with *Eco*RI was as described²¹ in the presence of 0.5 mM GpppG. Translation of proTGF- α mRNA was conducted as previously described^{22,23} in the presence of 50 μ Ci of ³⁵S-cysteine. Dog pancreas rough microsomal membrane vesicles²³ were added at a concentration of 0.05 equivalents μ l⁻¹. All translation mixtures were adjusted to 1.8 mM magnesium acetate and 85 mM acetate. After incubation for 80 min at 25 °C, the translation mixtures were centrifuged over a 0.5 M sucrose cushion. The membrane pellets, but not the supernatants unless otherwise indicated, were immunoprecipitated by overnight incubation at 4 °C with 1/30 dilution of anti-TGF- α or anti-proTGF- α rabbit antisera. The immunocomplexes were harvested with protein A-Sepharose. The anti-TGF- α antibodies⁹ were raised against a synthetic peptide corresponding to residues 71-88 of the rat proTGF- α sequence (residues 33-50 of mature rat TGF- α) and anti-proTGF- α antibodies⁹ were raised against a mixture of two synthetic peptides corresponding to residues 138-151 and 145-159 of rat proTGF- α , respectively. Endoglycosidase H digestions were performed according to supplier's specifications (Boehringer Mannheim) on material eluted from protein A-Sepharose beads by boiling in the presence of 50 mM Tris-HCl, pH 6.8, 50 mM dithiothreitol, 2% SDS.

† To whom correspondence should be addressed.

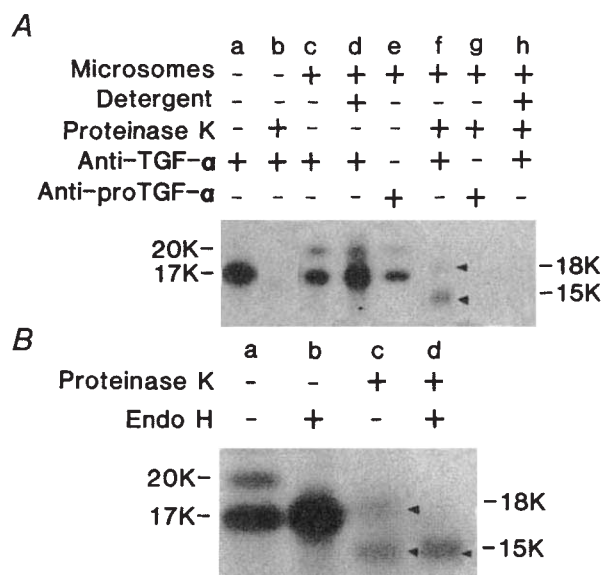


Fig. 2 Sensitivity of proTGF- α mRNA translation products to proteolytic digestion. **A**, Products of translations in the absence or presence of microsomes were incubated for 1 h on ice with (lanes b, f-h) or without (lanes a, c-e) 100 $\mu\text{g ml}^{-1}$ of proteinase K (Boehringer Mannheim). Some samples (d, h) also contained 0.5% of the detergent Nonidet P-40. The proteolytic digestions were arrested by adjustment to 2 mM phenylmethylsulphonylfluoride. The products immunoprecipitated from these samples by anti-TGF- α or anti-proTGF- α antibodies were displayed by SDS-PAGE and fluorography. Arrowheads, 18K and 15K products generated by proteinase K. **B**, Translation products obtained in the presence of microsomes, incubated with proteinase K or buffer alone and immunoprecipitated with antibodies to TGF- α . Some of the samples were then treated with endoglycosidase H as in Fig. 1. They were then displayed by SDS-PAGE and fluorography.

in many secretory and membrane proteins¹⁰, and core oligosaccharyl transferase which transfers high-mannose oligosaccharides to selected asparagine residues¹¹. The amino-acid sequence predicted for proTGF- α ⁶ includes an amino-terminal stretch of 23 uncharged and predominantly hydrophobic amino acids which resembles a typical signal sequence¹², followed by a potential *N*-linked glycosylation site at Asp 25 (ref. 6). Digestion of the translation products with endoglycosidase H was used to determine whether the microsome-associated 20K product corresponded to a cotranslationally glycosylated form of proTGF- α . Treatment of immunoprecipitated translation products with endoglycosidase H eliminated the 20K product (Fig. 1B, lanes e-f). No forms of proTGF- α with M_r smaller than the 17K product were observed in endoglycosidase-treated samples from translations containing microsomal membranes. These results imply that the signal sequence was not removed from either form of proTGF- α . Endoglycosidase digestion of products from translation reactions lacking microsomal membranes did not alter the M_r of the 17K product (Fig. 1B, lanes a-c).

Next, we examined whether the membrane association of the 20K product and the 17K product with microsomes was due to complete translocation of these polypeptides into the lumen as secretory molecules, or to their integration as membrane proteins. Membrane integration of proTGF- α could occur if the highly hydrophobic amino-acid sequence located on the carboxy-terminal side of the mature TGF- α sequence acted as a 'stop transfer' signal to leave proTGF- α anchored as an integral membrane protein. Two approaches were used to examine this possibility. The first approach was based on protease-protection assays. If proTGF- α were fully translocated across the microsomal membrane as a secretory protein it should be completely resistant to digestion by exogenous proteases. In contrast, if proTGF- α were a transmembrane protein, only those protein domains which were located inside the microsomal vesicle should be protected from proteolysis, whereas cytoplasmically exposed protein domains would be digested. Immunoprecipitation with sequence-specific antibodies should reveal which protein domains are within the lumen of the vesicles. Microsomal vesicles containing the 17K and 20K translation products were treated with proteinase K. This treatment resulted in the disappearance of both the 17K and 20K polypeptides and the generation of 15K and 18K products (Fig. 2A, lanes c and f). These 15K and 18K products were recognized by anti-TGF- α antibodies but not by anti-proTGF- α antibodies (Fig. 2A, lanes f, g). Treatment of the samples with endoglycosidase H after they

had been exposed to proteinase K showed that the 18K proteolytic product was the glycosylated form of the 15K proteolytic product (Fig. 2B, lanes, c, d). If microsomal vesicles were not present during translation (Fig. 2A, lanes a, b) or were disrupted with detergent after translation (Fig. 2A, lanes d, h), the 17K and 20K products were no longer protected from complete degradation by proteinase K. Thus, the 15K protease-resistant fragment must be derived from the membrane-integrated non-glycosylated 17K form of proTGF- α .

The other approach used to determine the membrane-spanning nature of proTGF- α takes advantage of the fact that exposure to alkaline conditions opens the microsomal vesicles and releases extrinsic membrane proteins as well as vesicle contents¹³. Treatment of microsomes with alkali after translation, followed by centrifugation yielded a pellet which retained a significant proportion of the 17K product as well as most of the 20K product (Fig. 3, lanes c, d). These products seem to correspond to the non-glycosylated and the glycosylated forms of proTGF- α , respectively, inserted into the microsomal membranes. That portion of 17K product that appeared in the control (Fig. 3, lane b) as well as alkali-treated supernatants most probably corresponds to proTGF- α that failed to be directed to the

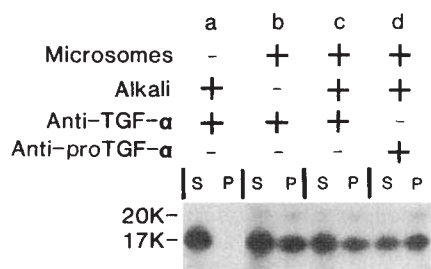
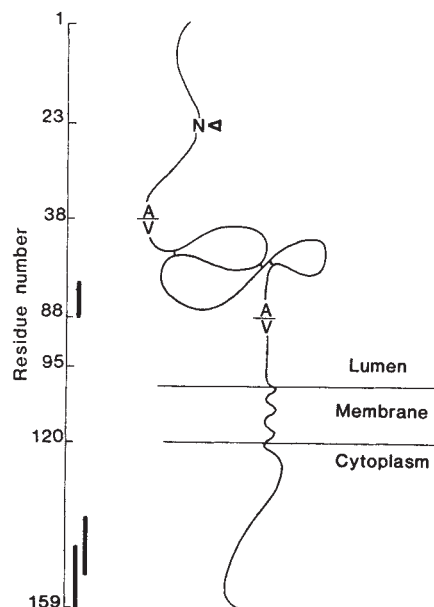


Fig. 3 Resistance of microsome-associated translation products to alkali extraction. ProTGF- α mRNA was translated in the absence (A, B) or presence (C, D) of microsomes. After translation, the samples were diluted fivefold with 0.1 M Na_2CO_3 , pH 11.5. After 30 min at 0 °C, the samples were layered on top of a cushion of 0.2 M sucrose (50 μl) containing 100 mM Na_2CO_3 , pH 11.5, 85 mM potassium acetate, 1.8 mM magnesium acetate, and centrifuged for 8 min in a Beckman Airfuge A-100/30 rotor at 30 p.s.i. The resulting pellets (P) and supernatants (S) were neutralized with acetic acid, subjected to immunoprecipitation with either anti-TGF- α or anti-proTGF- α antibodies, and analysed by SDS-PAGE and fluorography.

Fig. 4 Membrane topology of proTGF- α . Rat proTGF- α is a 159-amino-acid polypeptide⁶ that spans the membrane of the rough endoplasmic reticulum once, with the hydrophobic sequence between amino acids 97 and 120 acting as a membrane-spanning domain. The amino-terminal domain of the proTGF- α faces the luminal side of the endoplasmic reticulum. This domain contains the 50-amino-acid mature TGF- α sequence flanked by alanine-valine (A/V) sequences that act as cleavage sites for elastase-like enzymes (ref. 9). This domain also contains at its amino-terminal end a 23-amino-acid hydrophobic signal sequence and a *N*-linked glycosylation site at residue 25 (NA). The carboxy-terminal domain of proTGF- α faces the cytoplasmic side of the endoplasmic reticulum. Bars, sequences corresponding to the synthetic peptides used for generation of anti-TGF- α antibodies (residues 71–88) and anti-proTGF- α antibodies (residues 138–151 and 145–159) used in these studies. This model also applies to human proTGF- α ⁵, a 160-amino-acid polypeptide that exhibits 93% amino-acid sequence identity to rat proTGF- α , and has a hydrophobic sequence between residues 98 and 121 identical to the transmembrane sequence rat proTGF- α .



microsomes in this translation system. However, under all conditions tested the 20K glycosylated proTGF- α was associated almost exclusively with the microsomal membranes (Fig. 3, lanes c, d).

These results are consistent with the topological properties of proTGF- α illustrated in Fig. 4. According to the results, proTGF- α is integrated into the endoplasmic reticulum membrane and glycosylated at Asp 25. The hydrophobic sequence between residues 97 and 120 acts as a membrane-spanning domain that leaves the amino-terminal domain containing the mature TGF- α sequence (residues 39–88) located in the lumen of the endoplasmic reticulum, and the carboxy-terminal domain (residues 121–159) exposed on the cytoplasmic face of the membrane. The mature TGF- α sequence is flanked in proTGF- α by alanine and valine-rich sequences^{5,6} which can act as cleavage sites for elastase-like enzymes. Elastase does indeed convert TGF- α precursors into the mature product *in vitro*⁹. The subcellular location of the endogenous elastase-like enzyme that catalyses prohormone conversion in TGF- α -producing cells is unknown. However, the activity of the proTGF- α converting enzyme(s) is limiting in certain transformed cells which release and accumulate in the medium 17–19K bioactive intermediate forms of TGF- α ⁹. These soluble TGF- α intermediate precursors correspond to the extracellular domain of proTGF- α containing *N*-linked and sialylated *O*-linked carbohydrate attached to the 11K polypeptide chain (J.T. and J.M., manuscript in preparation).

Both the 17K and 20K membrane-integrated forms of proTGF- α appear to retain the amino-terminal signal sequence as judged from their identical mobility (following endoglycosidase H treatment of the 20K form) with the 17K form

synthesized in the absence of microsomal membranes. Because an uncleaved amino-terminal signal sequence was unexpected, we examined the proTGF- α sequence preceding the glycosylation site at residue 25 for potential signal peptidase cleavage sites following the criteria of Von Heijne¹². A potential signal peptidase cleavage site was detected between residues 22 and 23 (Cys-Gln-Ala/Leu) based upon the occurrence of small neutral amino acids at positions –1 and –3 (ref. 12). The lack of signal sequence cleavage in proTGF- α is presumably due to the short distance between the potential Ala/Leu cleavage site and the hydrophobic core of the signal sequence which extends up to residue 19 (ref. 14).

Cleavage of the extracellular domain of a membrane protein to yield hormonally active polypeptides as it occurs in the conversion of proTGF- α into its soluble bioactive products is a novel mode of prohormone processing. However, this mechanism of hormone release may also operate in the generation of bioactive polypeptides from the epidermal growth factor precursor^{15,16}, the vaccinia virus growth factor^{17,18} and the transcript from the neurogenic *Notch* locus in *Drosophila*¹⁹. All these proteins contain regions with amino-acid sequence identity to mature TGF- α , followed by a hydrophobic sequence that may act as a membrane-spanning domain as it does in proTGF- α . Indeed, antibodies against the EGF precursor stain the surface of cells transfected with an EGF-precursors expression plasmid²⁰. It is intriguing that a mechanism for generation of TGF- α /EGF-like intercellular signalling polypeptides by cleavage from membrane-bound precursors may have been conserved for proteins encoded by viral, insect and mammalian genomes.

This research was supported by the American Cancer Society by a grant to J.M., and by NIH grants to R.G. and D.C.L.

Received 18 December 1986; accepted 26 February 1987.

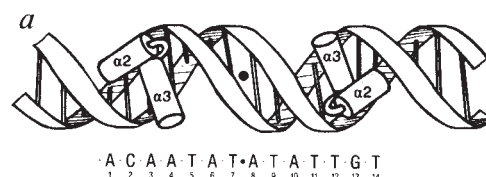
- Marquardt, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4684–4688 (1983).
- Massagué, J. *J. biol. Chem.* **258**, 13607–13613 (1983).
- Todaro, G. J., Fryling, C. & DeLarco, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5258–5262 (1980).
- Twardzik, D. R. *Cancer Res.* **45**, 5413–5416 (1980).
- Derynck, R., Roberts, A. B., Winkler, M. E., Chen, E. Y. & Goeddel, D. V. *Cell* **38**, 287–297 (1984).
- Lee, D. C., Rose, T. M., Webb, N. R. & Todaro, G. J. *Nature* **313**, 489–491 (1985).
- Blöbel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496–1500 (1980).
- Sabatini, D. D., Kreibich, G., Morimoto, T. & Adesnik, M. *J. Cell Biol.* **92**, 1–22 (1982).
- Ignatz, R. A., Kelly, B., Davis, R. J. & Massagué, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6307–6311 (1986).
- Evans, E. A., Gilmore, R. & Blöbel, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 581–585 (1986).

- Rothman, J. E. & Lodish, H. F. *Nature* **269**, 775–780 (1977).
- Von Heijne, G. *Eur. J. Biochem.* **133**, 17–21 (1983).
- Fujiki, Y., Fowler, S., Shio, H., Hubbard, A. & Lazarov, P. *J. Cell Biol.* **93**, 103–119 (1982).
- Von Heijne, G. *J. molec. Biol.* **173**, 243–251 (1984).
- Gray, A., Dull, T. J. & Ullrich, A. *Nature* **303**, 722–725 (1983).
- Scott, J. *et al. Science* **221**, 236–240 (1983).
- Blomquist, M. C., Hunt, L. T. & Barker, W. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7363–7367 (1984).
- Stroobant, P. *Cell* **42**, 383–393 (1985).
- Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. *Cell* **43**, 567–581 (1985).
- Bell, G. I. *et al. Nucleic Acids Res.* **14**, 8427–8446 (1986).
- Connolly, T. & Gilmore, R. *J. Cell Biol.* **103**, 2253–2261 (1986).
- Erickson, A. H. & Blöbel, G. *Meih. Enzym.* **96**, 84–93 (1983).
- Walter, P., Ibrahim, I. & Blöbel, G. *J. Cell Biol.* **91**, 545–550 (1981).

Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro

Gerald B. Koudelka, Stephen C. Harrison & Mark Ptashne

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA



The repressor of phage 434 binds to six operator sites on the phage chromosome¹⁻³. A comparison of the sequences of these 14-base-pair (bp) operator sites³ reveals a striking pattern: at five of the six sites, the symmetrically arrayed outer eight base pairs (four in each half-site) are identical and the remaining site differs at only one position (Fig. 1b). In contrast, the sequences of the inner four base pairs are highly variable. Crystallographic analysis of the repressor-operator complex⁴ shows that at each half-site, the 'recognition α -helix' of the repressor is positioned in the major groove such that it could contact the outermost five base pairs, but not the innermost two (Fig. 1a). We show in this paper that the sequence of the central base pairs of the operator (two in each half-site) have a significant role in determining operator affinity for repressor, despite the evidence presented here and in the accompanying paper⁴ that these base pairs are not contacted by repressor. We also show that these central base pairs influence operator affinity for Cro, a second gene regulatory protein encoded by phage 434 (ref. 5). We discuss the likely structural basis for this evidently indirect, but sequence-dependent, effect of the central base pairs of the operator on its affinity for the two regulatory proteins.

For the experiments presented here we synthesized a series of mutant 434 operators. Each of these operators differs from the twofold symmetric reference '14mer' at a single position, or at two symmetrically related positions in the central four base pairs (Fig. 1b). We studied the affinities of these operators *in vitro* for three proteins: 434 repressor, the amino-terminal DNA-binding domain of repressor (R1-69) and 434 Cro.

Table 1 shows that changing positions 7 and 8 (the two central base pairs) together from T·A-A·T to A·T-T·A had no effect on the affinity of the operator for repressor. In contrast, a coordinated change to either G·C-C·G or C·G-G·C decreased the affinity of the operator for repressor 50-fold. Similar results were obtained for operators that had changes at positions 6 and 9: a double substitution of A·T-T·A with T·A-A·T had no effect on the affinity of the operator for repressor, whereas both double changes, A·T-T·A → G·C-C·G and A·T-T·A → C·G-G·C at these positions decreased binding fivefold.

A single (asymmetric) T·A → G·C change at position 7 had a smaller (six- to sevenfold) negative effect on the affinity of the operator for repressor than did the double T·A-A·T → G·C-C·G changes at positions 7 and 8 (Table 1). A single T·A → A·T change at position 7 had a small (threefold) positive effect on the affinity of the operator for repressor (Table 1).

Substituting the central base pairs also affects the operator's affinity for both Cro and R1-69 (Table 2). Comparing Tables 1 and 2 reveals that the patterns of the affinity changes observed in response to the various substitutions were similar for all three proteins. Note also that any given base-pair substitution had a larger effect on the operator's affinity for the intact repressor than for either R1-69 or Cro.

Inosine, like guanine, base pairs with cytosine, but, like adenine, it lacks an N2 amino group that could protrude into the minor groove. Replacing the T·A-A·T base pairs at positions 7 and 8 with I·C base pairs has no effect on the operator's affinity for repressor, R1-69, or Cro (Tables 1 and 2).

434 Operator sequences

Position: 1 2 3 4 5 6 7 8 9 10 11 12 13 14

Naturally occurring operators

O_{R1} A C A A G A A A G T T T G T
O_{R2} A C A A G A T A C A T T G T
O_{R3} A C A A G A A A A A C T T G T
O_{L1} A C A A G G A A G A T T G T
O_{L2} A C A A T A A A T A T T G T
O_{L3} A C A A T G G A G T T T G T

Synthetic operators

14mer A C A A T A T A T A T T G T
6T A C A A T T T A A A T T G T
6C A C A A T C T A A G A T T G T
6G A C A A T G T A A C A T T G T
6G/2 A C A A T G T A T A T T G T
7A A C A A T A A T T A T T G T
7A/2 A C A A T A A A T A T T G T
7C A C A A T A C A T A T T G T
7G A C A A T A G A T A T T G T
7G/2 A C A A T A G A T A T T G T
7I A C A A T A I C A T A T T G T

Fig. 1 a, Schematic representation of the 434 repressor-operator complex. The helix-turn-helix structures of the dimer are shown in the configuration they have when bound to the synthetic operator⁴. Below, sequence of the operator illustrating the position of each operator base pair with respect to the helix-turn-helix structure. b, Sequences of the naturally occurring and synthetic 434 operators along one DNA strand. ●, Centres of symmetry of the operators. The circled base in the sequence of O_{R3} denotes the only base in the symmetrically arrayed outer eight bases of the naturally occurring operators that is not completely conserved. The bases in the synthetic operators that differ from the reference 14mer sequence are boxed.

Does repressor come into contact with the central base pairs? An examination of Fig. 1a suggests that a protein-DNA contact might be made on either the front side of the operator in the minor groove or on the back side in the major groove. The repressor of phage λ exploits the latter strategy, using a flexible 'arm' to make a sequence-specific contact to the major groove functional groups of the bases near the centre of its operator^{6,7}. 434 repressor, however, does not have an amino-terminal arm^{4,8}. The structural analysis of the repressor-operator complex⁴ indicates that Arg 43 could approach bases in the minor groove in the front of the operator. To test whether this residue might allow 434 repressor to discriminate between A·T and G·C base pairs at positions 6-9, we changed Arg 43 to Ala by directed mutagenesis and studied the binding *in vitro*, of the mutant protein to the reference and mutant operators. This protein, like wild-type repressor, bound more tightly to the reference operator than to operators bearing G·C substitutions at positions 6, 7 and 8 (Table 3), although its affinity for the reference operator was decreased about 500-fold. Because the mutant protein bound to operator only at concentrations near its non-specific binding affinity, a precise value of its discrimination

Table 1 Affinity of 434 repressor for synthetic operators

Operator	Relative affinity
14mer	1
6T	1.5
6C	5
6G	5
6G/2	2.5
7A	1
7A/2	0.3
7C	50
7G	50
7G/2	7
7I	1

For nomenclature used for operators, see Fig. 1. Binding affinities are expressed as the concentration of repressor monomers needed to occupy half-maximally each operator in a DNase I protection experiment. These values have been normalized to the amount of repressor needed to half-maximally occupy the reference operator. This value was determined in an experiment performed in parallel. Under the conditions used in these experiments 1 corresponds to 2×10^{-8} M.

Methods. The operators were synthesized on an Applied Biosystems model 380A DNA synthesizer and were purified by gel electrophoresis. The isolated single strands were annealed and ligated¹⁵ into the unique *Sal*I site in pUC18 (ref. 16). Except for the 7I operator, the ligation products were introduced into *Escherichia coli* strain JM101 (ref. 17). Plasmid DNA of resulting ampicillin-resistant colonies was screened for presence of the operator by restriction-enzyme analysis. The sequence of the operator was confirmed by DNA sequencing¹⁸. The operator-containing *Eco*RI-*Hind*III fragment, 3'-end-labelled at the *Eco*RI end, was purified by gel electrophoresis¹⁵. The 7I operator-containing *Eco*RI-*Hind*III fragment was obtained directly from the ligation products, without prior transformation. DNase I protection experiments were performed essentially as in ref. 19. The DNase I protection assay conditions were: 10 mM Tris, pH 7.8, 50 mM KCl, 1 mM MgCl₂, 2.5 µg ml⁻¹ sonicated chick blood DNA, 100 µg ml⁻¹ bovine serum albumin, 1 nM operator DNA and repressor at concentrations which varied by increments of 1.5-fold within each titration. Following incubation at 25 °C for 15 min, sufficient DNase I was added to give, on average, one cleavage per DNA molecule in 5 min of further incubation.

between A·T and G·C base pairs at positions 7 and 8 could not be determined, but it is at least fivefold. The mutant protein discriminates between G·C and A·T at position 6 as well as the wild-type does (2.5-fold). The mutant protein also distinguishes between the reference operator and other higher affinity operators as well as wild-type repressor does (data not shown).

Thus, two lines of evidence suggest that the central four base pairs in the 434 operator affect its affinity for repressor by influencing DNA conformation and not by interacting directly with the protein. First, operators that have either T·A-A·T or A·T-T·A base pairs at the symmetric operator positions 7 and 8 or at 6 and 9 bind equally well to repressor, whereas operators that have either G·C or C·G base pairs at these positions bind repressor equally poorly. We know of no other set of operator mutations where the pattern A·T = T·A > G·C = C·G has been observed. For example, this pattern of changes in the affinity of the operator for repressor differs from that observed when base substitutions are made at the outer five base pairs in each half-site. Repressor prefers one base at those positions, binding poorly to operators that have any of the three mutant bases (unpublished results). The structure of the R1-69 operator complex, described in the accompanying paper, shows that all the outer position bases can be contacted directly by side chains of the recognition α -helix⁴. Second, the crystallographic investigation⁴ shows that there is only one amino acid of repressor, centred just above the minor groove, that could contact the central four bases pairs; mutating this residue, Arg 43, has no effect on the ability of the protein to distinguish between A·T and G·C base pairs at the centre of the operator.

Table 2 Affinities of 434 R1-69 and Cro for synthetic operators

Operator	R1-69	Cro
14mer	1	1
6T	1.3	0.6
6C	3	2.7
6G	3	2.7
6G/2	1.8	ND
7A	1.3	1.5
7A/2	0.6	0.6
7C	15	17
7G	16	17
7G/2	4	3.2
7I	1	1

ND, not determined. The binding affinities of the proteins are expressed as in Table 1. Under the conditions of these experiments (see Table 1 legend), the concentrations of proteins required to occupy half-maximally the reference operator were: for R1-69, 5×10^{-7} M; for Cro, 2×10^{-7} M.

Table 3 Affinity of repressor(Arg 43 → Ala) for synthetic operators

Operator	Relative affinity
14mer	1
6G/2	2.5
7G	>5
7G/2	>5

Binding affinities of the 434 repressor(Arg 43 → Ala) are expressed as in Table 1. The conditions of these experiments were the same as described in legend to Table 1 except that the KCl concentration was 10 mM. Site-directed mutagenesis of 434 repressor was performed essentially as described in ref. 20. The mutant protein, bearing the Arg43 → Ala substitution, was purified following ref. 21. The concentration of this protein needed to occupy the reference operator half-maximally was 1.2×10^{-5} M.

The six amino-acid residues at the carboxy terminus of R1-69 (residues 64-69) are unresolved in the structure of the protein-DNA complex, but the model suggests that no contacts between these residues and the central bases can be made⁴. Residues 64-69 cannot contact these bases in the minor groove because access is blocked by residues 40-44. Access to the central bases in the major groove is possible on the back side of the operator. However, model building suggests that even in an extended conformation, residues 64-69 cannot reasonably reach around DNA to make contact with the central bases in the major groove. In experiments not shown, we found that an intact mutant repressor, bearing alanine in the place of arginine at position 69, bound the various operators listed in Fig. 1b identically to wild-type repressor. Moreover, inosine, which differs from guanine only in the minor groove, behaves like adenine, not guanine, when placed in the centre of the operator (Tables 1 and 2). This result makes it highly unlikely that a contact in the major groove is responsible for the observed effects. Finally, consistent with the idea that contact is made in neither the minor nor the major groove near the centre of the operator, repressor does not protect from methylation either adenine or guanine bases that are placed in various combinations at the central positions and methylation of these bases does not block repressor binding (ref. 2 and unpublished results).

Although we do not yet know the structure of the Cro-DNA complex, the experiments of Wharton *et al.*² and the high degree of amino-acid sequence homology between Cro and repressor⁸ suggest that it is very similar to the repressor-DNA complex. Hence, the central base pairs might also affect operator affinity

for Cro by influencing DNA conformation. We note that the amino-acid sequence of Cro differs from that of repressor at all positions homologous to repressor residues 64–69 (ref. 8). This difference suggests that these amino acids are not responsible for the observed effects on operator affinity for either Cro or repressor.

The operator in the 434 R1-69-DNA complex is overwound and slightly bent near its centre, and the minor groove of this region is compressed⁴. Evidently, G·C base pairs at positions 6–9 adversely affect binding of the protein by making the overwinding and bending of the DNA more difficult. This idea is consistent with the proposals of Drew *et al.*^{9,10}, who argued from the analysis of nuclease digestion experiments that the minor groove of A+T rich DNA is more readily overwound and compressed than is that of G+C-rich DNA. Drew *et al.* also concluded that minor grooves of poly(A) tracts are particularly susceptible to compression. This susceptibility could explain why an operator with the single T·A→A·T change at position 7, which creates a sequence of three A residues, has an increased affinity for protein. Application of Calladine's rules^{11,12}, which have been used to predict DNA structure, cannot account for the effect of G·C substitution at position 6–9 on the affinity of the operator for the proteins. In their simplest form, these rules do not distinguish between the purines; hence they would predict that operators containing G·C bases at positions 6–9 should have the same DNA structure (and by hypothesis, the same affinity) as operators containing A·T bases at these positions.

Removal of the N2 amino group of guanine, by substituting inosine, eliminates the deleterious effect of this base in our experiments. We cannot account for the negative effect of guanine by simply assuming that this N2 amino group clashes with the ring of an adjacent purine on the opposite strand: in the operator 7G/2 there is no purine positioned to make such a clash; nevertheless this operator binds the proteins less well than does the reference 14mer. We therefore imagine that the negative effect of G·C base pairs is the result, at least in part, of the base-pair hydrogen bond the N2 amino group forms with cytosine. The extra hydrogen bond could make the G·C base pair less deformable and therefore increase the energy required to form the altered DNA structure observed in the R1-69-DNA complex.

The DNA at operator positions 6–9 is configured to bring the half-sites of the operator into proper alignment with the protein, thus allowing each monomer of the bound dimer to make optimal contacts within each operator half-site⁴. Although the crystal structure reveals interactions between the two DNA-bound amino-terminal domains, biochemical experiments show that the intact repressor dimer is held together primarily by interactions between the two carboxy-terminal domains¹³. We therefore suppose that a dimer of two amino-terminal domains is more flexible than one of two intact monomers. This flexibility could explain why substituting G·C for A·T base pairs at positions 6–9 affects operator affinity for R1-69 less than its affinity for intact repressor.

The 434 repressor and Cro bind to the six naturally occurring operator sites with different orders of affinity^{2,3} and these differences are crucial to the life cycle of the phage¹⁴. Inspection of Fig. 1b reveals that two sites, *O*_R2 and *O*_L1, differ only in the central base pairs. Repressor and Cro both bind more tightly to *O*_L1 than to *O*_R2 (ref. 3) and evidently this discrimination is determined entirely by non-contacted bases. Although we do not have a complete description of how these proteins differentiate among the six naturally occurring operators, we believe that the indirect effect of the identity of the central base pairs has an important function in distinguishing the sites.

We thank members of the Ptashne and Harrison laboratories, especially Robin Wharton, Ed Giniger, Cynthia Wolberger and John Anderson for materials, discussions and for providing data in advance of publication. G.B.K. is supported by a Fellowship

of the Damon Runyon-Walter Winchell Cancer Fund. This work was supported by a grant from NIH to M.P. and S.C.H.

Received 8 January; accepted 1 April 1987.

1. Johnson, A. D. *et al.* *Nature* **294**, 217–223 (1981).
2. Wharton, R. P., Brown, E. L. & Ptashne, M. *Cell* **38**, 361–369 (1984).
3. Wharton, R. P. thesis, Harvard Univ. (1985).
4. Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 888–891 (1987).
5. Grosschedl, R. & Schwarz, E. *Nucleic Acids Res.* **6**, 867–881 (1979).
6. Eliason, J. L., Weiss, M. A. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2339–2343 (1984).
7. Pabo, C. O., Krovatin, W., Jeffrey, A. & Sauer, R. T. *Nature* **298**, 441–443 (1982).
8. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. *Nature* **298**, 447–451 (1982).
9. Drew, H. R. & Travers, A. A. *J. molec. Biol.* **186**, 773–790 (1985).
10. Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191**, 659–675 (1986).
11. Calladine, C. R. *J. molec. Biol.* **343**–352 (1982).
12. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257**, 14686–14707 (1982).
13. Anderson, J. E. thesis, Harvard Univ. (1984).
14. Ptashne, M. *A Genetic Switch* (Blackwell, Palo Alto, 1986).
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
16. Yannisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103–119 (1985).
17. Messing, J. *Meth. Enzym.* **101**, 29–79 (1983).
18. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
19. Johnson, A. D., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5061–5065 (1979).
20. Zoller, M. J. & Smith, M. *Nucleic Acids Res.* **10**, 6487–6500 (1982).
21. Anderson, J. A., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307–1311 (1984).

A new-specificity mutant of 434 repressor that defines an amino acid–base pair contact

Robin P. Wharton* & Mark Ptashne

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

The repressor encoded by bacteriophage 434 binds to its operators by inserting a 'recognition' α -helix into the major groove of the DNA^{1,2}. We have identified an amino acid–base pair contact that determines (in part) the DNA-binding specificity of 434 repressor. The identification is based on the properties of a 'new-specificity' mutant, named Repressor[Ala 28], which bears the substitution of Ala for Gln at the first residue of its recognition α -helix. Repressor[Ala 28] binds with high affinity to a particular doubly mutant operator bearing the same substitution at position 1 in each half-site, but does not bind to either the wild-type operator or to other mutant operators. We describe molecular models of residue 28–base pair 1 interactions that account for the binding specificities of both the mutant and wild-type proteins.

As each recognition helix in a repressor dimer interacts with one operator half-site, we imagined that a new-specificity mutant protein would recognize a doubly mutant operator bearing the analogous substitution in each half-site (see Fig. 1). To isolate new-specificity mutants, therefore, we first synthesized a set of twelve doubly mutant operators as well as a reference 'wild-type' site named the 14mer. Each of the mutant operators has one of the three possible base-pair substitutions (with reference to the 14mer) at position 1, 2, 3 or 4 in each half of the operator. (For example, the '1T' operator has a mutant T·A base pair at position 1 in each half-site, as shown in Fig. 2.) None of the mutant sites binds wild-type repressor detectably *in vitro* (data not shown).

To provide a rapid assay for the binding of candidate new-specificity mutant repressors *in vivo*, we modified the *Escherichia coli lacZ* gene so that the wild-type *lac* operator could be replaced (in separate bacterial strains) with each of the thirteen

* Present address: Howard Hughes Medical Institute, Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, 722 West 168 Street, New York, New York 10032, USA.

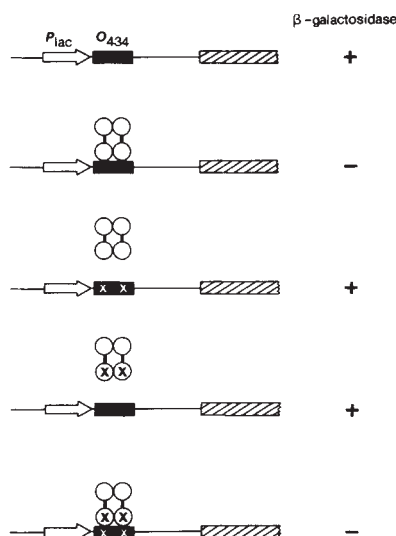


Fig. 1 Schematic drawing of the experimental plan for isolating new-specificity mutant 434 repressors. Each line in the left column represents the modified *lacZ* gene present in a different bacterial cell. Each cell contains, in addition, wild-type repressor (represented by a dimer of dumbbells on lines 2 and 3), or a new-specificity mutant repressor (represented by a dimer of dumbbells marked with black crosses on lines 4 and 5), or no repressor (line 1). Each repressor monomer is drawn as a dumbbell to indicate that it consists of two domains connected by a protease-sensitive linker. Transcription of each of the modified *lacZ* genes is directed by the wild-type *lac* operon promoter (arrow) which is regulated by the occupancy of either a wild-type or doubly mutant 434 operator (represented by a black box and a black box with two crosses, respectively). The doubly mutant operator bears the same substitution at the analogous position in each half-site. In the right column, '+' indicates that β -galactosidase is expressed at the fully depressed level and '-' indicates that its expression is at least partially repressed. Youderian *et al.*¹⁷ have described a similar experimental plan.

synthetic 434 operators described above (Fig. 2). Wild-type 434 repressor repressed the synthesis of β -galactosidase in the bacterial 'tester' strain bearing the 14mer but did not do so in any of the tester strains bearing mutant derivatives of the 14mer (data not shown). The extent of repression as determined by measuring β -galactosidase levels in permeabilized cells was only about twofold, but this degree of repression was sufficient to cause colonies to appear moderately red rather than dark red on MacConkey lactose agar plates³.

We then used directed mutagenesis to generate a set of eight genes, each encoding a repressor bearing a different amino acid at the first residue of its recognition helix (residue 28). In these mutant proteins, either Ala, Leu, Ser, Met, Arg, His, Lys or Pro was substituted for Gln at residue 28. Plasmids encoding each mutant repressor were introduced into the twelve tester strains bearing mutant operators as well as into the strain bearing the 14mer to yield 104 different transformants. When grown on MacConkey lactose indicator plates, only one of these 104 transformants—the one producing Repressor[Ala 28] and bearing a 1T mutant operator—formed light-red colonies, indicating that β -galactosidase synthesis was being repressed; the other 103 transformants formed dark-red colonies. Measurement of β -galactosidase levels showed that Repressor[Ala 28] repressed β -galactosidase synthesis approximately threefold in the 1T tester strain and did not repress β -galactosidase synthesis significantly in any of the other strains (data not shown).

In DNaseI protection experiments, we found that purified Repressor[Ala 28] half-occupied 1T mutant operators at a con-

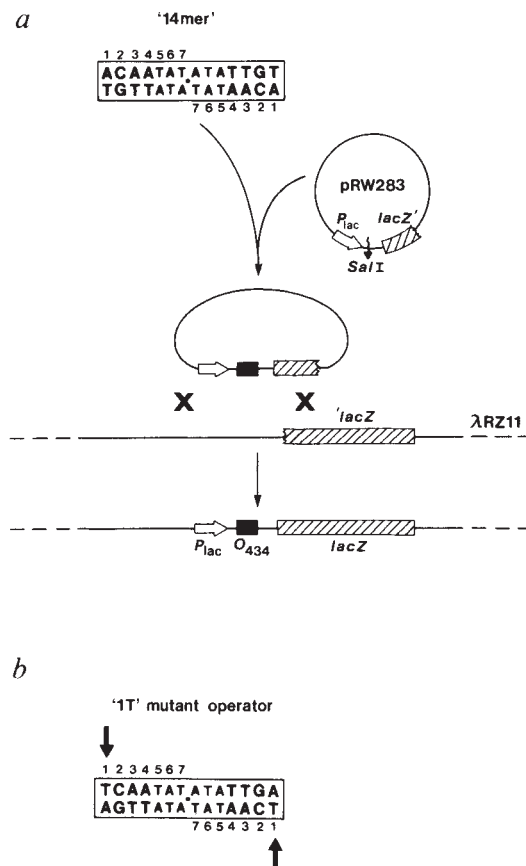
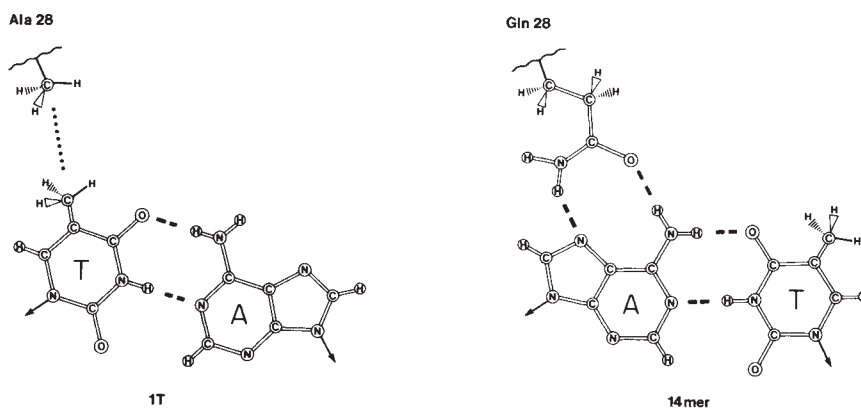


Fig. 2 Construction of bacterial 'tester' strains bearing synthetic 434 operators. **A**, At the top is a representation of the 14mer, with positions in each half-site numbered according to the convention used in the text. The synthetic 14mer and the wild-type site, O_{12} , differ at a single position, O_{12} bearing an A-T base pair at position 7 in one half-site. The remainder of the drawing (not to scale) shows schematically the construction of a bacterial tester strain in which expression of the *lacZ* gene is under the control of the 14mer. First, a synthetic duplex containing the 14mer was inserted into pRW283 at its unique *SalI* site, generating the unnamed intermediate plasmid in the middle of the drawing. Next, this intermediate was recombined *in vivo* (denoted with black crosses) with bacteriophage λ RZ11 (ref. 18) to generate the phage derivative shown at the bottom of the drawing. On this derivative, the 14mer (denoted O_{434}) sits between the *lac* operon promoter and an intact *lacZ* gene, as indicated. In an additional step not shown in the drawing, a single copy of the λ RZ11 derivative is stably integrated into the genome of an appropriate bacterial host (as a λ prophage), thereby generating the 14mer tester strain (see ref. 18 for details). A similar technique has been described by Benson *et al.*¹⁹. The *SalI* site on pRW283, which was introduced by directed mutagenesis, is at the start of the transcript synthesized from the *lac* operon promoter (P_{lac}). The *lacZ'* gene on pRW283 encodes the α -complementing fragment of β -galactosidase. The '*lacZ*' gene on λ RZ11 does not have a promoter and thus provides no β -galactosidase to cells carrying it. The bacterial host strain in these experiments was NK5031 (F^- *lacZ* Δ MM5265 *suIII*⁺ *Nal*^R) strain collection. **B**, A representation of the structure of the '1T' operator is shown as an example of the symmetric, doubly mutant derivatives of the 14mer used in screening for new-specificity mutants. The arrows indicate that the 1T operator is identical to the 14mer except at position 1 in each half-site, where the 1T operator has a T-A base pair and the 14mer has an A-T base pair. The '1C' operator has the sequence CCAATATATATTGG, the '2A' site has the sequence AAAATATATATTGT, and so on.

Fig. 3 Models of the interactions between residue 28 and base pair 1 for Repressor[Ala 28] and wild-type repressor. The drawing (to scale) on the left shows the Ala 28 side chain of Repressor[Ala 28] making a van der Waals contact (denoted by the dotted line) to the thymine 5-methyl group of the T·A base pair at position 1 in the 1T operator. The drawing on the right (also to scale) shows the Gln 28 side chain of wild-type repressor making a pair of hydrogen bonds (denoted by dashes) to the Adenine N7 and N6 atoms of the A·T base pair at position 1 in the 14mer. In each drawing, the remainder of the protein is indicated by a wavy line and the N-glycosyl bonds that attach base pair 1 to the rest of the operator are indicated by arrows. On the left, the distance between the methyl group carbon atoms is about 4 Å and on the right the distance between hydrogen bond donor and acceptor atoms is about 3 Å. We arbitrarily represented the hydrogen bond made by the Gln 28 carbonyl as linear and that made by the amide as somewhat non-linear (see Seeman *et al.*⁵). The relative positions of the protein and DNA are the same in each drawing. From these relative positions, Ala 28 cannot make a comparable van der Waals contact with any base pair other than T·A, and Gln 28 cannot make a comparable pair of hydrogen bonds with any base pair other than A·T. The drawings were adapted from drawings of ref. 5.



centration of 3×10^{-8} M, but it did not bind detectably to either the 14-mer or to 1A, 1C, 1G, 2A, 2G or 2T mutant operators at a 50-fold higher concentration (Table 1). Under the same reaction conditions, wild-type repressor half-occupied 14-mer operators at a concentration of 1×10^{-8} M and it did not bind detectably to 1T, 1C, 1G, 2A, 2G or 2C mutant operators at a 150-fold higher concentration.

The results described above strongly suggest that residue 28 of repressor interacts with base pair 1 in the operator, a conclusion consistent with that based on analysis of the crystal structure of the 434 repressor-operator complex⁴. We constructed molecular models of residue 28-base pair 1 interactions for both Repressor[Ala 28] and wild-type repressor. As shown in Fig. 3, the Ala 28 methyl side chain can be placed so that it specifically recognizes a T·A base pair by making a van der Waals contact to the 5-methyl group of the thymine. From the same relative positions of protein and DNA backbones, the Gln 28 side chain of wild-type repressor can be placed so that it specifically recognizes an A·T base pair by making a pair of hydrogen bonds to the adenine N6 and N7 atoms. This mode of adenine-specific recognition has been suggested by Seeman *et al.*⁵.

To test the idea that the thymine 5-methyl group at base pair 1 is important for operator recognition by Repressor[Ala 28], we measured its affinity *in vitro* for two mutant operators bearing unconventional bases at position 1. The 1U (for uracil) operator is a derivative of the 1T operator in which the pyrimidine at base pair 1 lacks a 5-methyl group, and the 1-5MeC (for 5-methyl cytosine) operator is a derivative of the 1C operator in which the pyrimidine at base pair 1 has a 5-methyl group. Repressor[Ala 28] did not bind detectably to 1U operators at a concentration 50-fold higher than that needed to half-occupy 1T operators (Table 1). Repressor[Ala 28] bound to 1-5MeC operators whereas it did not bind detectably to 1C operators; however, its affinity for 1-5MeC operators was about 18-fold lower than that for 1T operators. Wild-type repressor did not bind detectably to either 1U or 1-5MeC operators (Table 1).

Our model for the mechanism by which Repressor[Ala 28] recognizes the 1T operator implies that van der Waals interactions can be important in determining binding specificity. Because Repressor[Ala 28] binds almost as tightly to the 1T operator as wild-type repressor binds to the 14mer, we imagine that the energy of the van der Waals contact between methyl groups in the former case is similar to the energy of the pair of hydrogen bonds in the latter. This suggestion agrees with previous estimates for the energy of each type of interaction—1.5–

3.4 kcal mol⁻¹ for van der Waals contacts^{6,7} and 0.5–1.5 kcal mol⁻¹ per hydrogen bond⁸. Protein-DNA van der Waals contacts in the major groove are thought to be important in operator recognition by several repressor proteins—434 repressor⁴, *E. coli trp* repressor⁹, the bacteriophage λ repressor¹⁰ and λ Cro^{11–13}. In contrast, the restriction endonuclease *EcoRI* is thought to recognize its cleavage site by making only hydrogen bonds to the edges of base pairs¹⁴.

We do not understand why Repressor[Ala 28] binds so much more weakly to the 1-5MeC operator than to the 1T operator, because each site has a 5-methyl group on the appropriate base

Table 1 Wild-type repressor and Repressor[Ala 28] binding *in vitro*

Operator	Protein	
	Repressor	Repressor[Ala 28]
14mer	1	>150
1T	>150	3
1C	>150	>150
1G	>150	>150
1U	>150	>150
1-5MeC	>150	54

Each entry in the table corresponds to the relative concentration of either wild-type repressor or Repressor[Ala 28] at which the relevant operator was half-occupied in DNaseI protection experiments. For names of different operators, and repressor proteins, see text. The value '1' corresponds to a concentration of 1×10^{-8} M. Each entry of '>150' indicates that no specific binding was observed at a concentration corresponding to the value 150 (1.5×10^{-6} M), and that at slightly higher protein concentrations the entire operator-bearing DNA fragment was nonspecifically protected from cleavage with DNaseI. Not shown are the values for the 2A, 2G and 2T mutant sites, each of which is '>150'.

Methods. The generation of genes encoding mutant repressors and the preparation of mutant repressors to ~95% purity have been described¹⁶. DNaseI protection experiments were performed essentially as described² using sufficient DNase to cleave the labelled strand once (on average) in 1–5 min. The reaction conditions were: 10 mM HEPES (pH 7.0), 50 mM KCl, 1 mM MgCl₂, 0.2 mM CaCl₂, 2.5 μ g ml⁻¹ sonicated chick blood DNA, and 100 μ g ml⁻¹ bovine serum albumin (BSA). In experiments using the 1U or 1-5MeC operators, synthetic oligonucleotides were phosphorylated with T4 DNA kinase and ATP, self-annealed and ligated to *SalI*-cleaved pUC18. The resulting ligation products were cleaved with *EcoRI* and *PvuII*, and the *EcoRI* ends were labelled with [α -³²P]dNTPs and the Klenow fragment of DNA polymerase I. Operator-bearing fragments were purified by electrophoresis through a 40-cm long 7.5% polyacrylamide gel. The affinity of repressor for synthetic 14mer operators was identical to its affinity for cloned 14mer operators.

at position 1. Perhaps the relevant 5-methyl groups occupy different positions in space in the two operators as a consequence of slight changes in the DNA helix structure on substitution of a 5-MeC-G base pair for a T-A base pair. Because the energies of van der Waals interactions exhibit a very strong dependence on the distance between contacting atoms, a small difference in the relative atomic positions of the methyl groups could account for the large difference that we observe between the affinities of the 1T and 1-5MeC operators for Repressor[Ala 28]. Changes in DNA helix structure are thought to account for the effects on repressor binding of some base-pair substitutions at the centre of the 434 operator¹⁵.

We thank John Anderson and Steven Harrison for helpful discussions, Scot Munroe for gifts of plasmids, William Reznikoff for a gift of λ RZ11, and Betsy Burkhardt for artwork. This work was supported by an NIH grant to M.P. and Steven Harrison.

Received 24 March; accepted 26 March 1987.

1. Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **316**, 596-601 (1985).
2. Wharton, R. P. & Ptashne, M. *Nature* **316**, 601-605 (1985).
3. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
4. Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
5. Seeman, N. C., Rosenberg, J. M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 804-808 (1976).
6. Fersht, A. R., Shindler, J. S. & Tsui, W.-C. *Biochemistry* **19**, 5520-5524 (1980).
7. Goeddel, D. V., Yansura, D. G. & Caruthers, M. H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3578-3582 (1978).
8. Fersht, A. R. *et al.* *Nature* **314**, 235-238 (1985).
9. Schevitz, R. W. *et al.* *Nature* **317**, 782-786 (1985).
10. Lewis, M. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **47**, 435-440 (1983).
11. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. *Nature* **298**, 718-723 (1982).
12. Hochschild, A., Douhan, J. III & Ptashne, M. *Cell* **47**, 807-816 (1986).
13. Hochschild, A. & Ptashne, M. *Cell* **44**, 925-933 (1986).
14. McClarin, J. A. *et al.* *Science* **234**, 1526-1541 (1986).
15. Koudelka, G. B., Harrison, S. C. & Ptashne, M. *Nature* **326**, 886-881 (1987).
16. Wharton, R. P. thesis, Harvard Univ. (1985).
17. Youderian, P., Vershon, A., Bouvier, S., Sauer, R. T. & Susskind, M. M. *Cell* **35**, 777-783 (1983).
18. Yu, X.-M. & Reznikoff, W. S. *Nucleic Acids Res.* **12**, 5449-5464 (1984).
19. Benson, N., Sugiono, P., Bass, S., Mendelman, L. V. & Youderian, P. *Genetics* **114**, 1-14 (1986).

A novel type of energy metabolism involving fermentation of inorganic sulphur compounds

Friedhelm Bak & Heribert Cypionka

Fakultät für Biologie, Universität Konstanz, Postfach 5560, D-7750 Konstanz, FRG

Two processes are known whereby energy is conserved during substrate metabolism in heterotrophic organisms: respiration and fermentation. Both involve oxidation-reduction reactions; but whereas in respiration the electrons are transferred from substrate to an electron acceptor, in fermentation part of the substrate molecule itself accepts the electrons. Fermentation is therefore a type of disproportionation, and does not involve an overall change in oxidation state of the substrate. All fermentative substrates known to date are organic molecules. We have discovered a novel type of fermentation involving the disproportionation of inorganic sulphur compounds in certain sulphate-reducing bacteria¹. Initially discovered in a newly isolated sulphate-reducing bacterium, *Desulfovibrio sulfodismutans*, the capacity for disproportionation of sulphur compounds is also found in some known sulphate-reducing bacteria and various bacteria isolated from freshwater, brackish or marine sediments.

From anoxic sludge of a ditch near Konstanz a vibrio-shaped bacterium, strain ThAc01 (Fig. 1a), was isolated in pure culture. It grew in defined anoxic medium² with acetate and thiosulphate or sulphite, but not with acetate and sulphate. Surprisingly, the

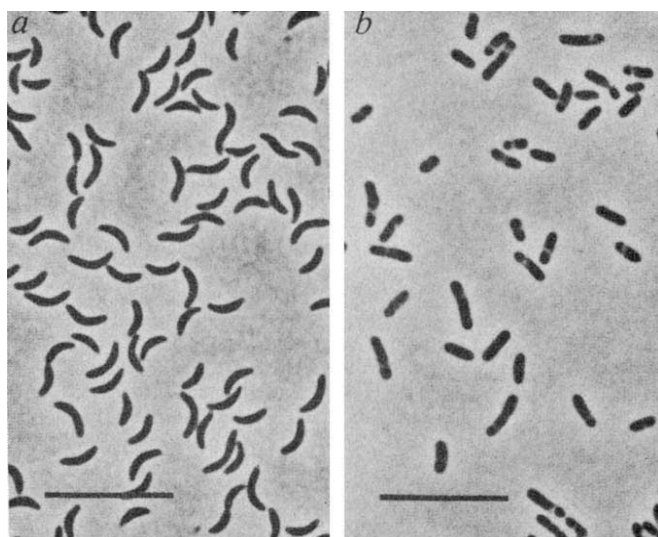


Fig. 1 Phase-contrast photomicrographs of *a*, *Desulfovibrio sulfodismutans* strain ThAc01, grown with sulphite as energy source, and with acetate and CO₂ as carbon sources; and *b*, strain NTA3 grown with thiosulphate as energy source and CO₂ as the only carbon source. Bar, 10 μ m.

strain grew also in medium with sulphite or thiosulphate, if no organic substrate was added. It was discovered that, under strictly anoxic conditions, not only sulphide but considerable amounts of sulphate had been produced. In the absence of cells neither sulphate nor sulphide was formed from sulphite or thiosulphate. Acetate was only required in low concentrations (1 mM) as carbon source. Growth yields were proportional to the amounts of thiosulphate or sulphite added, but were independent of the acetate concentration (1-10 mM). Determination of the stoichiometry (Table 1) revealed that 0.75 mol sulphate and 0.25 mol sulphide were produced per mol of sulphite; from thiosulphate, equal amounts of sulphate and sulphide were formed. The stoichiometries thus corresponded to the disproportionations given in equations (1) and (2) in Table 2. The electron balances found do not allow for a significant oxidation of acetate, which would have resulted in the formation of more sulphide and less sulphate.

Disproportionation of thiosulphate and sulphite represents a novel type of energy metabolism. In contrast to all biological transformations of sulphur compounds known so far, sulphite and thiosulphate are not unidirectionally oxidized or reduced, but only dismutated. Such reactions are known from inorganic sulphur chemistry, but spontaneously occur only at temperatures above 80 °C^{3,4}. Formally, the described metabolism resembles the fermentation of organic compounds, and may be termed inorganic or chemolithotrophic fermentation.

With some substrates other than acetate (lactate, ethanol, propanol or butanol), strain ThAc01 grew also by dissimilatory sulphate reduction, producing only sulphide from sulphate. The strain is named *Desulfovibrio sulfodismutans*, and is described extensively elsewhere⁵.

Bacteria able to disproportionate inorganic sulphur compounds appear to be widely distributed in nature. Out of ten known species of sulphate-reducing bacteria tested, two disproportionated sulphite, and one strain could also disproportionate thiosulphate. Sulphite- and thiosulphate-disproportionating bacteria were also isolated from various brackish water or marine sediments. Strain NTA 3 (Fig. 1b), isolated from anoxic sludge from Norsminde Fjord (Denmark), appeared even to be facultatively autotrophic. This bacterium grew by thiosulphate disproportionation over at least seven consecutive subcultures in defined medium with CO₂ as the only carbon source. In contrast,

Table 1 Stoichiometry of disproportionation of sulphite or thiosulphate by *Desulfovibrio sulfodismutans*

Sulphur compound consumed (mmol)	Sulphite	Thiosulphate
Sulphate formed (mmol)	11.5	9.1
Sulphide formed (mmol)	3.6	8.5
Ratio of sulphate formed to sulphide formed	3.19	1.07
Cell dry mass formed (g mol of sulphite or thiosulphate)	3.2	2.1

Cells were grown for 12 days at pH 7.2 and 35 °C in 1.2-litre bottles with defined mineral medium² in the presence of 2 mM acetate as carbon source. Sulphite was added in three incremental portions to avoid toxic effects. Sulphate, sulphite, thiosulphate, sulphide, cell dry mass and protein were determined as in ref. 8. Cell dry mass contained about 70% protein.

growth of *D. sulfodismutans* ceased after the third transfer in acetate-free medium. A most probable number (MPN) enumeration of thiosulphate-disproportionating bacteria with mud from the freshwater ditch near Konstanz, revealed high numbers of 2×10^6 sulphate- and sulphide-forming cells per ml mud.

The disproportionation of sulphite yields a free energy change of $-58.9 \text{ kJ mol}^{-1}$ per sulphite, the disproportionation of thiosulphate $-21.9 \text{ kJ mol}^{-1}$ (Table 2). Thus, sulphite disproportionation will not allow for the conservation of more than 1 ATP molecule per sulphite molecule, and the disproportionation of thiosulphate even less. The low growth yields of *D. sulfodismutans* (Table 1) are in agreement with this.

Further exergonic disproportionations to sulphate and sulphide are those of dithionite, or polythionates, whereas disproportionation of elemental sulphur is endergonic (equations (3) to (5)). *D. sulfodismutans* grew also by disproportionation of dithionite; however, because of its instability and the impurity of commercially available dithionite, the stoichiometry of dithionite disproportionation was not determined. The addition of tetrathionate to the culture resulted in the formation of elemental sulphur and sulphate; however, no growth was observed. Presumably, sulphide was produced, reacting chemically with tetrathionate to form elemental sulphur. Elemental sulphur was found to inhibit growth of *D. sulfodismutans*.

Other conceivable sulphite or thiosulphate disproportionations are those resulting in the formation of sulphate and elemental sulphur (equations (6) and (7)). These reactions would be accompanied by even greater changes of the free energy. However, formation of elemental sulphur was not observed in cultures of *D. sulfodismutans* supplied with sulphite or thiosulphate.

Table 2 Gibbs free energy changes of disproportionation for some inorganic sulphur compounds

Reaction	ΔG° (kJ mol ⁻¹) (at 25 °C and pH 7)
(1) Sulphite $\rightarrow$ sulphate + sulphide $4 \text{ SO}_3^{2-} + \text{H}^+ \rightarrow 3 \text{ SO}_4^{2-} + \text{HS}^-$	-235.6 (-58.9 per sulphite)
(2) Thiosulphate $\rightarrow$ sulphate + sulphide $\text{S}_2\text{O}_3^{2-} + \text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + \text{HS}^- + \text{H}^+$	-21.9
(3) Dithionite $\rightarrow$ sulphate + sulphide $4 \text{ S}_2\text{O}_4^{2-} + 4 \text{ H}_2\text{O} \rightarrow 5 \text{ SO}_4^{2-} + 3 \text{ HS}^- + 5 \text{ H}^+$	-536.0 (-134.0 per dithionite)
(4) Tetrathionate $\rightarrow$ sulphate + sulphide $4 \text{ S}_4\text{O}_6^{2-} + 12 \text{ H}_2\text{O} \rightarrow 9 \text{ SO}_4^{2-} + 7 \text{ HS}^- + 17 \text{ H}^+$	-360.2 (-90.0 per tetrathionate)
(5) Sulphur $\rightarrow$ sulphate + sulphide $4 \text{ S}^0 + 4 \text{ H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 3 \text{ HS}^- + 5 \text{ H}^+$	+40.9 (+10.2 per sulphur)
(6) Sulphite $\rightarrow$ sulphur + sulphate $3 \text{ SO}_3^{2-} + 2 \text{ H}^+ \rightarrow \text{S}^0 + 2 \text{ SO}_4^{2-} + \text{H}_2\text{O}$	-186.9 (-62.3 per sulphite)
(7) Thiosulphate $\rightarrow$ sulphur + sulphate $3 \text{ S}_2\text{O}_3^{2-} + 2 \text{ H}^+ \rightarrow 4 \text{ S}^0 + 2 \text{ SO}_4^{2-} + \text{H}_2\text{O}$	-106.5 (-35.5 per thiosulphate)

Free energy changes calculated after Thauer *et al.*⁹

The mechanisms of energy conservation during disproportionation of sulphite or thiosulphate have not been studied so far. There are at least two conceivable strategies of energy conservation. In the first, bacteria could utilize the ATP (or ADP) sulphurylase during sulphate formation. Usually, this enzyme acts in the reverse direction and catalyses the ATP-consuming activation of sulphate in sulphate-reducing bacteria. Such a process in *D. sulfodismutans* would be similar to reactions known from aerobic sulphur-oxidizing⁶ and from phototrophic sulphur bacteria⁷, and would represent a substrate-level phosphorylation of the sort typical of fermentations. Second, during disproportionation of sulphite, energy might be conserved through the building up of a chemiosmotic gradient coupled to sulphide formation. During disproportionation of thiosulphate, the oxidation states of the two sulphur atoms of the thiosulphate molecule are not changed; thus, electron transport phosphorylation is less probable in this case.

We thank Professor Dr N. Pfennig and Dr F. Widdel for valuable discussions. This work was supported by the Deutsche Forschungsgemeinschaft.

Received 5 December 1986; accepted 19 February 1987.

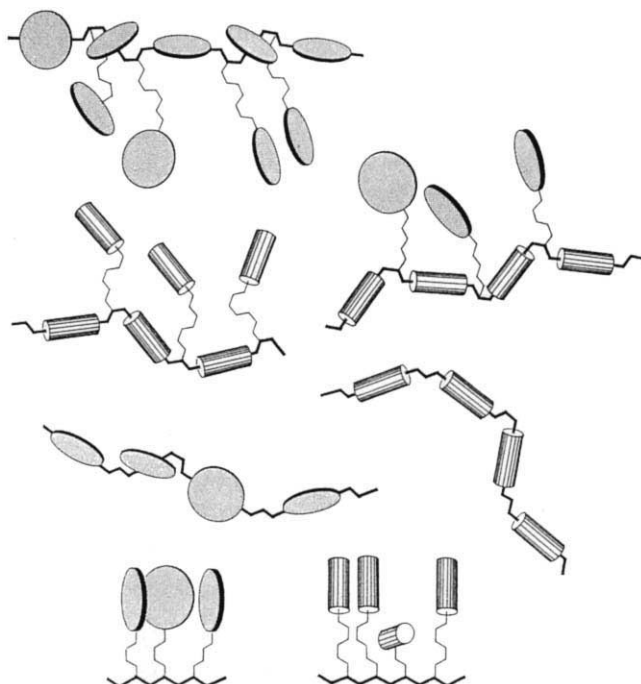
1. Postgate, J. R. *The Sulphate-reducing Bacteria* 2nd edn (Cambridge University Press, 1984).
2. Widdel, F. & Pfennig, N. *Arch. Microbiol.* **129**, 395-400 (1981).
3. Schmidt, M. in *Sulfur, its Significance for Chemistry, for the Geo-, Bio- and Technosphere and Technology* (eds Müller, A. & Krebs, B.) 259-276 (Elsevier, Amsterdam, 1984).
4. Belkin, S., Wirsen, C. O. & Jannasch, H. W. *Appl. Environ. Microbiol.* **49**, 1057-1061 (1985).
5. Bak, F. & Pfennig, N. *Arch. Microbiol.* (in the press).
6. Kelly, D. P. *Phil. Trans. R. Soc. P298*, 499-528 (1982).
7. Trüper, H. G. *Phil. Trans. R. Soc. B298*, 529-542 (1982).
8. Cypionka, H. & Pfennig, N. *Arch. Microbiol.* **143**, 396-399 (1986).
9. Thauer, R. K., Jungermann, K. & Decker, K. *Bact. Rev.* **41**, 100-180 (1977).

Erratum

Optical technology: Novel optical properties of liquid-crystal polymers

George Attard & Graham Williams
Nature **326**, 544-545 (1987).

IN this News and Views piece this figure was poorly printed and details of shading lost. It is a schematic representation of the rich molecular architecture of liquid-crystalline polymers, showing side-chain types with disk- and cigar-type moieties.



What's wrong with SDI?

Paul Doty

Star Wars in a Nuclear World. By Lord Zuckerman.
William Kimber, London: 1986. Pp.226. £12.50.

In a recent examination of 'Star Wars', Professor Harvey Brooks observed:

Considered as an example of American science policy, the Strategic Defense Initiative (SDI) seems virtually unique. There is no parallel from the rich and diverse menu of the past that can provide a plausible model for the probable evolution and fate of SDI as a piece of 'big science'.

The most expensive scientific and technical undertaking in history was launched in 1983 with almost no assessment, analysis or advice from those best placed to give it. As Lord Zuckerman points out, the White House Science Council and its chairman, who was science advisor to the President, were ignorant of the decision and the President's speech announcing it until five days before its delivery. The project manager of the main effort to be folded into SDI learned of the decision only on the day of the speech. Thus the Iranian 'missiles-for-hostages' caper was not the first failure of the advisory systems that were thought to guide Presidential decision-making in this administration.

It is therefore most fitting that the dean of British science advisors has set forth his views on this extraordinary and most ambitious enterprise. This he does in the longest of the nine essays that make up *Star Wars in a Nuclear World*. But the title of the book is nevertheless misleading because the other eight essays cover a much broader range. Three are concerned directly with the nuclear world; the remainder with the role of science advising in both Britain and the United States during the Second World War and especially in the post-War period, the 40 years in which Zuckerman has been a continuing presence in Whitehall.

In the book, Zuckerman's overarching concern is how to connect the essential knowledge of specialists to the scientifically uneducated political leadership and how to choose wisely what new technologies justify heavy government investment. Again and again he shows the importance of introducing the advice at the right time, the irrelevance of quantification when qualitative issues are dominant and the overriding nature of the

political process which advising serves. The essays are in the best tradition of the genre: historically informed, well researched and thought-out, pungently stated and delivered with style. I liked "The Politics of Outer Space" best. It adds much insight to what is in any case a remarkable book, and allows the author to share his pessimism over the future of



command technology, that is the institutionalization of technological change for state purposes.

The essay on Star Wars (Chapter 4) delineates clearly the main technical weaknesses of SDI. This is done mostly as a commentary and critique on the debate among experts in the United States in 1984 and 1985. The choice of experts is well informed, the presentation fair, the verdict on the ultimate benefit of the enterprise negative:

...if one were to regard SDI as a bargaining chip, one would also have to accept that the US will gain only if it throws it away. ... Even if SDI were to confound its critics and succeed in the sense that its separate components could be fitted together in a working system, the United States and the West would lose, not only because the USSR would have devised measures for defeating a space-based ballistic

missile defence, but because there are other ways than land-based ICBMs ... whereby the US could be threatened with nuclear devastation.

Still, important points are not covered. One is the economic costs. Even the research and development phase is destined to outrun the entire cost of the Apollo programme with its 15 missions. The cost of deploying any high performance system that can now be envisaged is near the trillion dollar level and its continuous maintenance and upgrading would be a crippling drain on the defence budget. The opportunity costs are in the same high range. For example, a disproportionately large fraction of the best scientists and engineers will be taken from tasks that

may be more productive, militarily or otherwise. Many projects will be put aside to meet the relentless demands of SDI. The spin-offs from such expenditure will be there, to be sure, but they will not compare to the technology that could be developed with the same funds funnelled directly to more useful ends — there is simply not much scope for application in civilian life of the technologies, such as heavy space lift and space-based proton accelerators, that SDI would develop. The vulnerability of whatever is deployed, and the relatively low cost of counter-measures, is touched upon but not developed in a sufficiently convincing manner.

Another item that is usually overlooked in evaluating strategic defences is their destabilizing effect both on East-West relations and in upsetting the military balance. The most obvious interpretation of the extreme opposition from the Soviet Union is that it wants to be spared the expense of competing in this new dimension of warfare and will see the failure to bargain away SDI as a strategy for disrupting its programme for economic reform. If a strategic defence system, even of limited pretension, were deployed on either side it would elicit an increase in strategic offence on the other, as Secretary Weinberger has reminded us, as well as a step-up in efforts to match or to counter such a system. Moreover, if at some distant time both sides possessed such systems the grounds for suspicions would be vastly multiplied over those with which we deal so imperfectly today. The claims of the effectiveness of the other side's system would be unconstrained because the capability of any system would remain unknown to either side since no realistic testing would be possible. The uncertainty in the military balance under these con-

ditions would dwarf the most extreme claims of imbalance that we hear today.

Zuckerman foresees the elements in the debate that are now becoming clear. One is the growing recognition that what is proposed is indeed a 'supersystem' in systems engineering jargon, as C. A. Zraket pointed out in a recent paper in *Science*; that is, an aggregate of systems with hundreds or thousands of nodes interconnected over huge geographical areas and too complicated, dynamic and interactive for its performance to be understood *ab initio* no matter how much analysis is performed on it. Another is the crucial role to be played by the SALT I Anti-Ballistic Missile (ABM) Treaty. The US administration's attempt to reinterpret the treaty so as to permit testing and development of space-based ABM systems, despite the opposite intent of the treaty, justifies the fear that the treaty will be broken as the first of a sequence that could go on to ravage the Outer Space Treaty and the Limited Test Ban Treaty if nuclear explosives are to be developed for use in space.

Finally, Zuckerman raises the crucial question of whether a democracy can sustain the heavy expenditure involved for the very long time required when the outcome, in contrast to the Apollo programme for example, is so vague and uncertain. Indeed, the most striking current development is the restructuring of SDI to concentrate on short-term demonstrations of a prosaic nature at the cost of reducing the effort to develop exotic systems. The clear reason for this is to wrest maximum funding from an increasingly reluctant Congress.

Unfortunately, the European reaction to SDI receives only scant treatment. For the most part, European governments have acquiesced and public opposition to SDI is largely uninformed. It is only at this late hour that one foreign minister (Herr Genscher of West Germany) has taken a stand against the plans to erode the ABM Treaty. All of this is the essence of science advising too. Even though strategic defence stands no chance of becoming an effective country-wide shield, the effort to develop it can have profound and far-reaching political consequences. Reading Zuckerman's book might persuade some who have stood aside to enter the fray. If not, they will have benefited from a remarkable tutorial on how science serves as the seed-bed of technology, and on the Anglo-American experience in choosing what technologies to integrate into their respective social, political and military structures. □

Paul Doty is Mallinckrodt Professor of Biochemistry at Harvard University, and Director Emeritus of the Center for Science and International Affairs, John F. Kennedy School of Government, Cambridge, Massachusetts 02138, USA.

In the corridors of power

Daniel J. Kevles

Science, Technology, and Public Policy. By Richard Barke. *CQ Press*, 1414 22 Street NW, Washington, DC 20037:1986. Pp.245. Pbk \$10.95.

IN THE United States today, at least 50 federal agencies deal, in one way or another, with science and technology, and technical issues have come to pervade American governmental policy. Richard Barke, a political scientist on the faculty of the University of Houston, rightly stresses that science and technology enter policy in two different ways. One is policy for science and technology — that is, policy to call forth new technologies and to advance scientific knowledge. The other is policy that uses them to achieve particular goals, for example, a clean environment. Federal science and technology policy in the former sense is a product of the Second World War, which, in its aftermath, yielded wide-ranging federal patronage of research and development (R&D) through agencies such as the National Science Foundation, the Atomic Energy Commission (most of whose functions have been taken over by the Department of Energy) and the Office of Naval Research. Currently, the richest and most powerful federal patron of R&D is the Defense Department, which accounts for about 70 per cent of federal funds spent for the purpose.

Federal policy that uses science and technology dates back well into the nineteenth century, when, for example, the management of western settlement drew in part upon geological surveys. It expanded in the twentieth century as the federal government began to exploit expert knowledge for the conservation of natural resources and the regulation of foods and drugs. What has made scientific knowledge currently so pervasive in federal policy-making is the passage in the past 20 years of laws to strengthen environmental protection, occupational health and safety, public health and medicine, and consumer protection. In 1985, 33 statutes required federal agencies to consider risk and ways to reduce it. Under the circumstances, R&D has become an instrument — in certain respects, a weapon — of policy. As regards environmental matters, for example, R&D makes regulation possible; its absence — as the Reagan administration recognized when, in its early days, it sought to cut the research budget of the Environmental Protection Agency — can undermine or destroy regulatory efforts.

Yet, despite the double significance of

science and technology in federal affairs, there is no single, overarching policy for them. (American policy for science and technology, it has been remarked, is to have no policy.) There is, rather, an array of policies that cover the multifarious technically related areas in which the federal government is engaged.

Barke argues that, for all this diversity, federal science and technology policies can be analysed with a certain degree of commonality because they are all subject to the constraints of law, knowledge, coordination and politics. He takes that analysis as his task here, while conceding that "devising a formal theory of public policy ... is analogous to the attempts of nuclear physicists to discover a grand unified theory of the four known forces of the universe" (p.153). Barke does not pretend to arrive at any conclusion about how particular technical policies — or, *a fortiori*, the aggregate of them — come to be made. But he does provide a very useful primer on the institutions, practices and problems of science and technology policy-making, and his treatment is salted with illuminating and thought-provoking insights.

Through the 1960s, the President and the federal bureaucracy held the upper hand in science and technology policy-making. The President not only enjoyed control of the budgetary agenda and centralized authority over his administration; he could also call upon a richly knowledgeable corps of technical advisors within the bureaucracy and could command the services of distinguished people outside of it, a power symbolized by the creation, in 1957, of the President's Science Advisory Committee (PSAC). Though PSAC has been transformed and demoted, the President retains the services of the bureaucracy, including the Office of Management and Budget, whose reach is immense and whose internal examiners include a number of PhDs in the sciences.

However, Congress is better equipped now than it was in the 20 years or so after the Second World War to check and balance the power of the White House. Both the House and the Senate have an extensive committee structure devoted to different aspects of science and technology, and defence R&D has become the special concern of subcommittees in the Armed Services Committees in the two houses. Congress can also turn to several omnibus support agencies, notably the Congressional Budget Office and the Office of Technology Assessment, which enjoys burgeoning Congressional confidence and which, while not really a match for the White House advisory apparatus, still provides valuable analyses and reports on diverse topics ranging from the effects of nuclear war to drug bioequivalence.

One of Barke's arresting observations concerns the considerable degree to which the courts have become involved in matters of science and technology policy. What brought the courts to prominence were the disputes, beginning in the 1960s, over the particular issue of nuclear power and the general one of environmental protection. The National Environmental Policy Act of 1969 gave citizens, interest groups and firms the right to block proposed federal actions, and, as Barke notes, helped lead to the filing before the end of the 1970s of nearly 1,200 lawsuits. Environmental measures, not to mention those addressed to occupational health

the trend bespeaks a clash over the goals of policy — whether what is to be sought is the efficient advancement of science or a geographically and institutionally more widespread scientific capability — that only politics can resolve.

Politics figures still more powerfully in policy addressed to what the physicist Alvin Weinberg has called 'trans-scientific' questions — that is, questions that go beyond the technical 'is' (or, more often, technical 'maybe') to human 'oughts'. Barke observes that "the task of risk management is necessarily political", adding, "Given the uncertainties of data and methods that plague risk analysis,

PURPOSE

SEC. 2. The purposes of this Act are: To declare a national policy which will encourage productive and enjoyable harmony between man and his environment; to promote efforts which will prevent or eliminate damage to the environment and biosphere and stimulate the health and welfare of man; to enrich the understanding of the ecological systems and natural resources important to the Nation; and to establish a Council on Environmental Quality.

Statement of intent — the opening paragraph of the National Environmental Policy Act of 1969. The Act was the stimulus for a rash of lawsuits in the following years.

and safety, consumer protection and the like, stimulated

questions about the meanings of laws, the reliability of technical evidence, and inconsistencies in government actions; so it has fallen largely upon judges to determine such questions as what levels or risks Congress implicitly designated as acceptable, or how much ambiguity in analysis was permissible [p.96].

Legislation aside, the explosion of information and the availability of new technologies has also drawn the courts into adjudicating such complex issues as the rights of the unborn and of the aged.

Yet the legislative and legal system is constrained by the limits of knowledge. Barke aptly observes: "Statutes and judges cannot successfully require what is physically impossible" (p.15). At the same time, knowledge cannot dictate science and technology policy. That is inevitably subject to considerations of customs, values, costs and interests, all of which express themselves through the political system — through politics.

Since 1945, it has been a cardinal point of American policy that the most efficient way to advance science and technology is to award R&D funds to the best scientists and engineers in the best universities. In operation, that policy has led to a concentration of federal money in certain institutions and geographical areas, with the result that the have-nots have mobilized in recent years to bypass the peer review process and obtain funds for research facilities directly through the political process. Frank Press, president of the National Academy of Sciences, has called the trend a type of "special-interest political favoritism that can only hurt the overall scientific endeavor". Perhaps, but

experts can usually be found to take opposite sides of an issue" (p.85). Then, too, the technical fact that living next to a nuclear power plant is less risky than living in cosmic-ray-rich Denver, Colorado, does not necessarily make the neighbours of the power plant any more comfortable. In one of his case-study chapters, Barke demonstrates that the administration of hazardous waste policy has made indelibly clear that scientific knowledge, however certain, cannot resolve contests of political and economic trade-offs. At the same time, growth in technical knowledge — in this case, better understanding of the ecological consequences of various waste-disposal methods — can well lead to changes in environmental requirements and creation of the political will to impose them.

To Barke's mind, proposals to improve science and technology policy-making through administrative manoeuvres (for example, creating a Department of Science), better educating the public on technical matters or removing it from politics, are not only unrealistic but also unwise. They ignore the distributive nature of power in the federal system, the absorption of different groups in the polity mainly in issues of immediate concern to themselves, and, above all, the irreducibly political nature of science and technology policy-making. The fact of the matter is that science and technology policy is no longer something special but has become a regular — if not altogether ordinary — participant in the rough and tumble of the process of governing the United States. □

Daniel J. Kevles, a historian, is at the California Institute of Technology, Pasadena, California 91125, USA.

Spring Books

The Effects of Low Temperatures on Biological Systems

Edited by **Brian W. Grout** and **G. John Morris**

Probable publication May 1987
512 pages illustrated
£45 net approx boards
0 7131 2893 3

The Ecology of Sex

P. Greenwood and **J. Adams**

Publication April 1987 80 pages
illustrated £5.95 net paper
0 7131 2934 4

Anatomy of Flowering Plants

An introduction to structure and development

Paula Rudall

Probable publication April 1987
88 pages illustrated
(with half-tones) £6.95 net paper
0 7131 2950 6

Enzymes in Industry and Medicine

New Studies in Biology
Gordon F. Bickerstaff

Publication February 1987
104 pages illustrated
(with half-tone) £4.95 net paper
0 7131 2935 2

Human Genetics and Medicine

New Studies in Biology
Sir Cyril A. Clarke

Publication February 1987
128 pages illustrated
(with half-tones) £4.95 net paper
0 7131 2944 1



Edward Arnold

41 Bedford Square, London WC1B 3DQ

Investigations into mass mortality

A. Hallam

Extinction. By Steven M. Stanley. W.H. Freeman: 1987. Pp.242. £15.95.*

A GREAT deal of excitement has been aroused in recent years by theories of extraterrestrially induced mass extinctions, and of extinction periodicity. Hitherto, however, the non-specialist has been frustrated by the lack of a popular but authoritative account of these ideas, putting mass extinctions into the broader context of geological and biological history. Such an account has now been provided by the distinguished palaeontologist Steven Stanley. It comes in the style of an extended article from *Scientific American*, and as one has come to expect of Stanley the text is lucid, concise and remarkably free of jargon. Quite subtle concepts are explained with refreshing simplicity, and the book is illustrated by a series of excellent colour photographs and diagrams.

Although there have been five major episodes of mass extinction of the world's biota in the past 600 million years, together with a larger number of minor episodes, the greatest recent interest has focused on the end of the Cretaceous — partly because of the demise of everyone's favourite creatures, the dinosaurs, and partly because of the much-discussed evidence of an asteroid impact. Accordingly, Stanley devotes most attention to this episode and provides an excellent, well-balanced and remarkably up-to-date description of research on the subject. He relegates the role of a sudden global catastrophe, resulting from the impact of an extraterrestrial body or spectacular volcanicity, to an insignificant one compared with the longer-term environmental dis-

* In the United States published as a volume in the Scientific American Library.

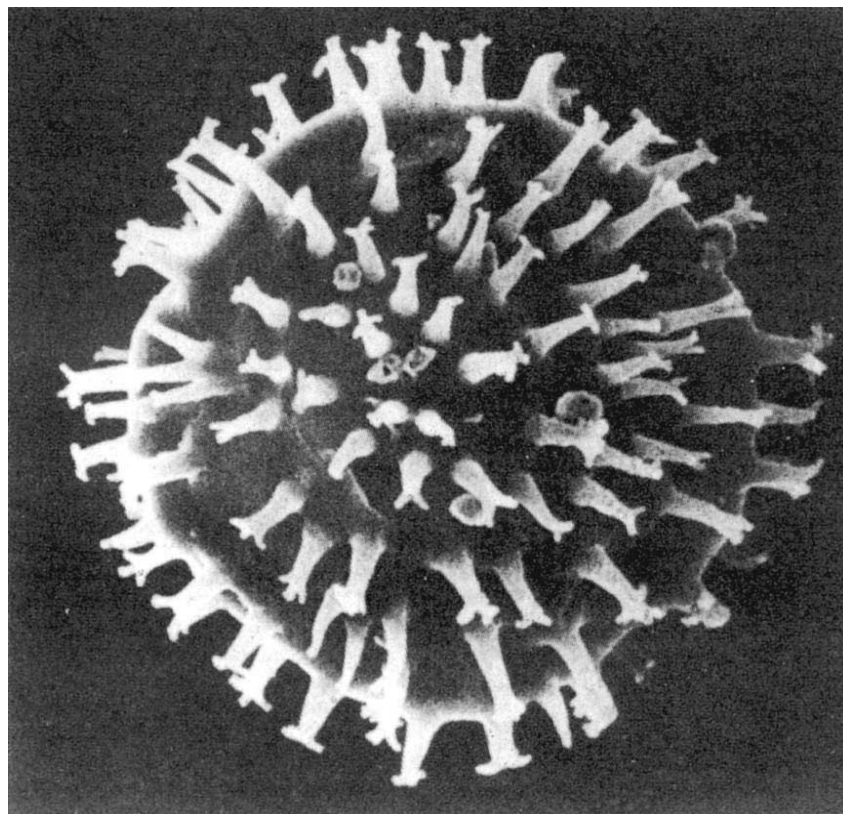
turbance produced by a global fall in temperature. Indeed, climatic deterioration as the main promoter of mass extinctions forms the leitmotiv of Stanley's book. In this respect he make a persuasive case for molluscan extinctions in the western Atlantic in the past few million years, and a reasonable one for increased rates of extinction among both marine and terrestrial organisms across the Eocene-Oligocene boundary.

Further back in time than the late Cretaceous, the evidence supporting his thesis is much more vague and equivocal. Stanley has in consequence to resort in varying degrees to special pleading. Nowhere is this more evident than in the part of the book dealing with the greatest mass extinction episode of all, at the end of the Palaeozoic era. Stanley constantly reiterates the point that tropical organisms such as reef corals have been the most vulnerable to extinction, but such organisms are likely to have exhibited a more general environmental sensitivity than merely to fall in temperature. Even for the apparently well-authenticated case of the Eocene-Oligocene boundary, the issue has been complicated by Shackleton's latest review of oxygen isotope data suggesting that the most severe episode of oceanic temperature fall took place *after* the extinction of most of the marine organisms in question. While Stanley is

right to cast doubt on the widely accepted view that loss of habitat area because of periodic falls of sea-level is a sufficient explanation of marine extinctions, he overstates his case by being selective and partial with the evidence. In particular he relies too uncritically on the sea-level curve of Vail and his associates, ignoring the fact that this curve has been the subject of much dispute. He also underestimates both the sharpness of many episodes of mass extinction as revealed by a close analysis of stratigraphic sections, and the role of anoxic bottom waters spreading episodically across epicontinental seas.

In the present inadequate state of knowledge, it seems that we should keep an open mind about a possible multiplicity of causes, taking due note of the fact that some of them may be intimately linked, such as sea level, climate and volcanicity. There is, after all, no good reason to assume that any two extinction episodes have had precisely the same cause. Perhaps, however, it is better that an author of a book intended for a wide public should enthusiastically promote one particular point of view rather than blandly sit on the fence. Stanley's work certainly makes for stimulating reading. □

A. Hallam is Lapworth Professor of Geology, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.



Victim of the first great extinction? An acritarch, the dormant stage of a single-celled alga which is well represented in the fossil record. The diversification of the acritarchs ended suddenly about 650 million years ago, coincidentally with widespread glaciation although the connection remains tentative.

From Extinction

Published April 1987

THE STUDY OF CULTURE: REVISED EDITION, by L. L. Langness • 288 pages, 6 x 9", paper, notes, references, index, illustrated, 1987, \$10.95, ISBN 0-88316-556-2

Study: Revised is an historically focused introduction to cultural anthropology with an emphasis on the concept of culture that will have strong interdisciplinary appeal. It represents both an expansion and updating of the first edition (1974), which has gone through eight printings and has been widely used in college and university courses.

L. L. Langness is Professor of Anthropology and Psychiatry at UCLA.

Order directly from Chandler & Sharp by sending check or money order (no credit cards) plus \$1.50 to cover postage and handling. California customers add appropriate sales tax.

Chandler and Sharp Publishers, Inc.
11A Commercial Boulevard
Novato, California 94947

The accounting for ourselves

Martin J.S. Rudwick

Theories of Human Evolution: A Century of Debate 1844–1944. By Peter J. Bowler. Basil Blackwell/Johns Hopkins University Press: 1987. Pp.318. £25, \$32.50.

CHARLES Darwin, in one of the most striking understatement in the history of science, noted at the end of the *Origin of Species* that through the further development of his theory "light will be thrown on the origin of man and his history". The problem of "man's place in nature", to use T.H. Huxley's phrase, was central to the intensive debates that followed the publication of the *Origin*, yet the fossil evidence for human evolution remained sparse and highly ambiguous well into the twentieth century. Although the history of palaeoanthropology has been recounted by some of its practitioners, it is only now that post-darwinian research into the origin of the human species has been analysed in detail by an historian of science. Peter Bowler is already well known for his books *Fossils and Progress* (Science History Publications, 1976) and *The Eclipse of Darwinism* (Johns Hopkins University Press, 1983); he has now put both historians and palaeoanthropologists in his debt by producing this fine summary of a hundred years of debate about human evolution.

The precise century specified in his subtitle need not be taken too seriously. Bowler really picks up the story at the point when a vast antiquity for the human species first came to be widely accepted, coincidentally in the years immediately before and after the publication of the *Origin* (1859); and his epilogue carries the story forward well beyond his nominal ending point, to include for example Johanson's 'Lucy' (*Australopithecus afarensis*) of 1974. That span of history includes the discovery of the Neanderthal fossils, Dubois' *Pithecanthropus* or 'Java Man', the notorious 'Piltdown Man' and the African australopithecines, to mention only a few. But this is no naively empiricist account of successive fossil finds. Partly as a result of the continuing scarcity of human and near-human fossils, the empirical evidence for human evolution was always open to a wide variety of interpretations; and those interpretations owed almost as much to broad assumptions about human nature as they did to the details of human anatomy. Yet conversely the interpretations were never entirely unconstrained by such evidence as was available; and since the evidence was by its nature cumulative — apart from the aberrant episode of the Piltdown

fraud — it can be seen to have exercised increasing constraints on the range of interpretations that could carry scientific conviction.

This feature of the debate would have emerged more clearly from Bowler's account, had he structured his book on more strictly narrative or chronological lines. As it is, his largely topical chapters tend to take the reader to and fro across the decades in a way that makes it difficult to sense the shifting course of opinions about the character of human evolution. He does however show in a striking way how non-darwinian assumptions of an essentially linear and intrinsically progressive trend in human evolution continued to pervade the theorizing of paleoanthropologists, even as late as the 1930s. But it is all the more surprising that he has chosen not to adopt a narrative mode for

his history, because he is evidently fascinated by Landau's and Cartmill's controversial suggestion that palaeoanthropological accounts of human origins themselves have an irreducibly narrative, and even mythic, structure.

In any case, Bowler himself emphasizes that his account is a preliminary one, and others who want to follow up his leads will be grateful to him for his thorough survey of a vast mass of relevant literature. Like his earlier books, this one is indeed mainly a survey of scientific literature rather than a history of the practice of science. Nonetheless, within the chosen limits of his work, Bowler has compiled a fine account of an evolutionary debate that touches us, perhaps, more nearly than any other. □

Martin J.S. Rudwick is Director of the Program in History of Science at Princeton University, Princeton, New Jersey 08544, USA.

Bones of history

Adrian Desmond

A Short History of Vertebrate Palaeontology. By Eric Buffetaut. Croom Helm/Methuen: 1987. Pp.223. £30, \$57.50.

HISTORY might not be bunk, but historians of science have had a hard time proving the point. Since Martin Rudwick published his subtle *Meaning of Fossils* in 1972, the antique bone-and-biography approaches have been superseded by a number of new contextual studies in geology and palaeontology. These offer a variety of cultural frameworks for appreciating the older interpretations of fossils — interpretations that are often at odds with our own in the late twentieth century.

Eric Buffetaut's book, by contrast, is a tunnel history of the more traditional type. It is written for working scientists, drawing largely upon palaeontological papers and older histories. For all that it is a sprightly and interesting work, especially in its latter part.

Much of the book deals with the formative decades of the nineteenth century, which began with Cuvier's discovery of the succession of fossil faunas. Already in that post-Revolutionary period, when 'revolutions' of the Earth were in vogue, good ichthyosaurs were fetching £150 and famous fossils such as the Stonesfield 'opossums' could cause academic brawls. Buffetaut discusses the chance discovery of many vertebrate fossils, often as a by-product of mining activities. It was perhaps the very serendipitous nature of the science that made it attractive to eccentric scholars, not a few of whom ended their careers with suicide (Hugh Miller, Vladimir Kovalevskii, Baron Nopcsa).

Buffetaut rightly hints that less impetus

was given to fossil studies by darwinism than by the contemporary imperial expansion. The colonists eventually cut down their supplies to Europe's capitals, diverting them to their own regional museums. And with the emergence of large-scale fossil expeditions at the end of the last century, these national museums grew to become themselves cultural showcases proclaiming colonial independence. Buffetaut ends his story at the First World War, when new phylogenetic goals and ecological reconstructions were paving the way for what seem so commonplace today: the museum dioramas.

A history of vertebrate palaeontology in English has long been a desideratum. And even though Buffetaut makes little concession to modern historiography and passes over many specialist historical studies, his *Short History* provides a baseline for further work. True, professional historians will be miffed at Buffetaut's Whiggish reconstructions — his back-dating of discoveries to accord with our modern understanding. (We have Robert Plot, unbeknown to himself, describing dinosaur bones in 1676.) They would also want to see the earlier so-called 'fanciful' theories (for example, explanations of fossils based on the scholastic notion of *vis plastica*) treated more sympathetically, perhaps using one of the sociological or philosophical approaches to science now on offer.

Nonetheless, Buffetaut has provided a neat synopsis of fossil discoveries worldwide. His book will prove useful to historians and will appeal to palaeontologists reared on Edwin Colbert's classic *Men and Dinosaurs*. □

Adrian Desmond is an Honorary Research Fellow in the Department of Zoology, University College London, Gower Street, London WC1E 6BT, UK, and author of *Archetypes and Ancestors: Palaeontology in Victorian London 1850–1875* (University of Chicago Press, 1984).

Earth science: time for reassessment

Gordon L. Herries Davies

Time's Arrow, Time's Cycle: Myth and Metaphor in the Discovery of Geological Time.
By Stephen Jay Gould. Harvard University Press: 1987. Pp.222. \$17.50, £14.25.

OUR author needs no introduction. Those who have enjoyed such earlier of Gould's books as *The Panda's Thumb* or *Hen's Teeth and Horse's Toes* will again find here the brand of entertainment that they relish. This new work arises from Gould's delivery of the first series of Harvard-Jerusalem lectures presented at Hebrew University in April 1985. It is a highly individualistic document (Gould admits it to be "a quest for personal understanding") and sometimes discursive (the book opens within the works of Sigmund Freud and closes outside the south front of the Cathedral of our Lady of Chartres), but it is always highly readable. To style it "stunning", as does the publisher on the dust-jacket blurb, is to indulge in hyperbole. But vastly entertaining and stimulating it most certainly is.

Gould's subject here is geological time; he is concerned with certain aspects of the discovery of what John McPhee has appropriately termed "deep time". The story of that discovery, from seventeenth-century exegesis to twentieth-century radiometric dating, is a tale oft told, and there is within Gould far too strong a pioneering instinct to allow him to rest content with the following of so well-trodden a trail. His is a more subtle approach.

He has given his book two overt themes. The main theme — whence the book derives its title — involves Gould's contention that as we in the Western world look back over the history of our globe we are torn by a conceptual dichotomy. On the one hand we believe in the arrow which moves steadily through time. Behind it there stretches a past now irredeemably lost to us; before it there lies a future which the arrow's continued flight steadily converts first into an instantaneous present and then secondly into the latest increment of an accumulating past. This concept of the arrow of time can hardly be unrelated to the presence within the Christian faith of a linear, directional view of history which traces events from a Creation, through a Fall, an Advent and a Crucifixion, to a future Parousia and a coming Millennium. But Gould is con-

vinced that mankind is less than happy with the arrow of time. It allows us to interpret history only as a series of unique events whereas "we also crave some underlying generality, some principles of order transcending the distinction of moments" (p. 196). The result of this craving is the interpretation of history not as a series of unique linear events but, rather, in terms of a series of continuously repeated cycles. This cyclic interpretation, one which seems to have found favour among most human societies, is the second element within Gould's dichotomy.

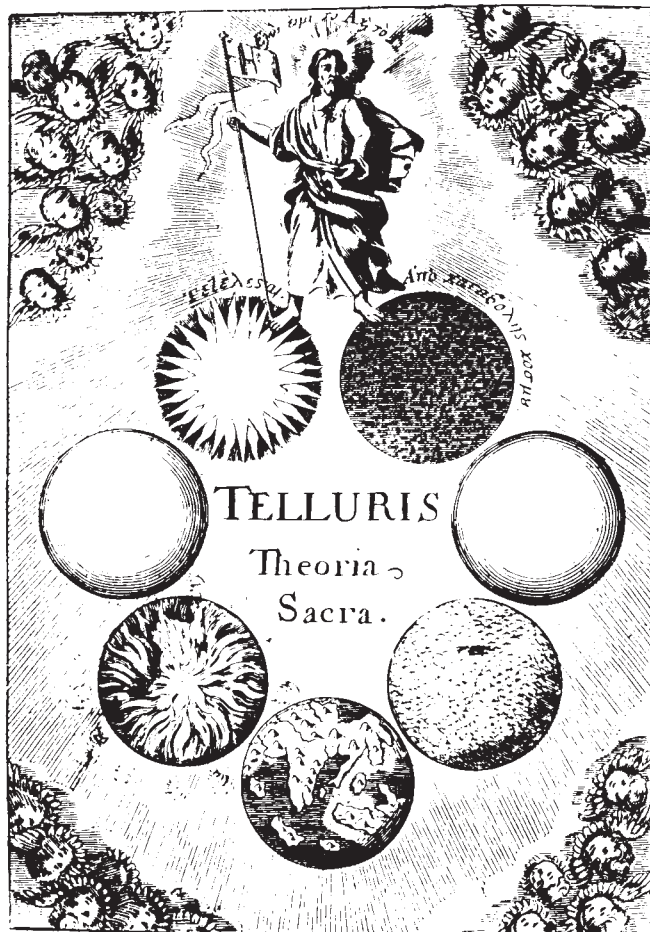
In the book Gould explores these two metaphors — the arrow and the cycle — in the context of geology's discovery of deep time, and he has clearly found satisfaction through his recognition of the dichotomy

between them. "For me", he writes (p.191), "this contrast became a key for unlocking both the structure and meaning of great historical documents that I had read several times before, but had never understood or grasped as unified statements". This, the first of Gould's two themes, is clearly of interest to all manner of intellectual historians.

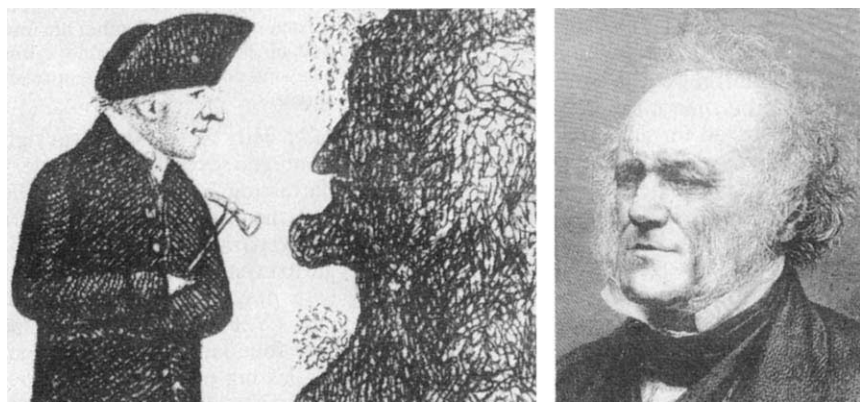
The second theme — the sub-plot — is directed expressly towards practising Earth scientists. Here Gould emphasizes for the benefit of all such readers that much of what passes for the history of the Earth sciences really bears scant relationship to the true passage of past events. Over the past 25 years a great deal of incisive scholarship has been directed towards the history of the Earth sciences, but as yet these studies have made little impact upon the historical understanding of the Earth sciences community at large. It is to be hoped that Gould's message will now get through to those quarters where it is most urgently needed. Those preliminary lectures — those textbook introductions — based upon nothing more substantial than Sir Archibald Geikie's *The Founders of Geology* of 1905 just will not suffice today. At best, such derivations from Geikie perpetuate simplistic half-truths; at the worst they perpetuate those popular myths which Gould brands as "textbook cardboard".

Gould here makes no attempt to demonstrate the validity of his theses through a comprehensive analysis of European literature at large. He prefers to adopt the case-studies approach, and offers a trio of refreshingly original essays devoted to the major publication of just three of the British scholars who loom large in virtually all our histories of the Earth sciences. The chosen three (one standard villain and two standard heroes, as Gould puts it) are Thomas Burnet as mirrored in his *Telluris Theoria Sacra* of 1681 and 1689, James Hutton as mirrored in his *Theory of the Earth* of 1795 and Sir Charles Lyell as mirrored in his *Principles of Geology* of 1830 to 1833.

Burnet and his theory have been harshly treated by a posterity which expects a history of the Earth to be based upon the study of subduction zones and unconformities rather than upon the Pentateuch and soaring flights of the human imagination. But Gould adopts a timely revisionist stance and he reminds us first that by the standards of his day — and what other standards may we justly employ? — Burnet was



From beginning to end — the frontispiece from Thomas Burnet's *Telluris Theoria Sacra* of 1681. The spheres represent the successive phases in the history of the Earth following a clockwise sequence from the top right as follows: (1) The chaotic liquid; (2) the Earth in its pristine state; (3) the Earth during the Flood; (4) the modern Earth; (5) the Earth during the conflagration which will precede the Parousia; (6) the Earth during the Millennium; and (7) the Earth in its ultimate state as a star. (Reproduced from *Time's Arrow, Time's Cycle*.)



Geological heroes — James Hutton (left) and Sir Charles Lyell.

a fine scholar, and secondly that throughout his work Burnet was at pains to demonstrate the efficacy of natural processes to account for all the postulated events of Earth history. More especially, he refused to invoke miraculous intervention. In this respect he is entitled to stand higher in our estimation than is Sir Isaac Newton because, in his correspondence with Burnet, Newton expressed the view that God might occasionally have interfered to change the Earth's speed of rotation.

It is upon the frontispiece to Burnet's *Theory* that Gould focuses particular attention and he sees it "as the finest expression ever published of the tension between two complementary views of time" (p.41). In many histories of the Earth sciences Burnet's whimsey is contrasted with the more securely field-based theory of the Earth published by Nicolaus Steno in 1669, but in a tailpiece to his essay Gould suggests that in reality both men were influenced by that same dichotomy between time's arrow and time's cycle.

In turning to James Hutton, Gould confronts one of the most revered figures in the history of the Earth sciences; in 1947 there was placed upon Hutton's Edinburgh grave a tablet proclaiming him to have been "The founder of modern geology". In this essay, Gould looses a flight of well-aimed bolts. First, he joins other recent authors in attacking the Geikie-derived myth that Hutton's theory of the Earth was the masterpiece of an empiricist deeply experienced in field phenomena. In reality, at the time of its formulation Hutton's theory was no more securely grounded upon observational evidence than was Burnet's, although it does need to be emphasized that the recently rediscovered illustrations that were to have accompanied the unpublished third volume of Hutton's *Theory* of 1795 reveal that late in his life Hutton did indeed become a true field geologist. Secondly, Gould argues that in Hutton's theory we see reflected a mind which has decisively rejected the arrow of time in favour of the cyclic view. I find this thesis persuasive. We must hope that the interpretation will

now be further explored in the context of the eighteenth-century deistical movement of which Hutton was a part. Thirdly, Gould observes that those who have criticized Hutton for his failure to enlist palaeontology and stratigraphy in the cause of his theory have misunderstood their man. Hutton, Gould claims, deliberately shunned fossils and the stratigraphic column because he had no concern for any kind of directionality in Earth history. He was interested solely in the Earth as a self-repairing, steady-state, perpetual-motion machine.

Gould's third essay — the longest and most closely reasoned — is devoted to Charles Lyell's brilliant *Principles*, and here he returns to a number of the points already made in connection with Hutton's theory. In Gould's own words, the object is to demonstrate that "Lyell was no impartial empiricist, but a partisan thinker committed to defending time's cycle against a literal record that spoke strongly against a directionless world..." (p.103). It is a thought-provoking piece of writing which will command the attention of all Lyell scholars.

Underlying the entire book, however, lurks yet another and still deeper theme which should commend the work to a readership far wider than historians of ideas and of science. Gould both explicitly and implicitly demonstrates that science is a creation of human minds which are ever feeling the influence of pressures far removed from those natural phenomena that are laid out before the scientist's gaze. Sir Peter Medawar has repeatedly reminded us that "The purely logical element in scientific discovery is a comparatively small one". Gould amply demonstrates precisely this point. Geology is not some mysterious miasma which gradually seeps from the Earth's surface to be collected and analysed by the Earth scientist. Rather is it a secretion which oozes generation by generation out of the human cerebrum. □

Gordon L. Herries Davies is Associate Professor of Geography at Trinity College, University of Dublin, Dublin 2, Ireland, and is General Editor of the Dictionary of Irish Biography.

SPRING BOOKS

FROM LONGMAN
SCIENTIFIC & TECHNICAL

THE DINOSAUR HERESIES

By Robert Bakker

Overwhelming evidence that, far from being a flawed dynasty of sluggish monsters plodding through the swampy Mesozoic terrain, the dinosaurs were a warm-blooded and highly successful species.

April ISBN 0 582 00420 9 Cased £14.95

SECRETS OF SLEEP

By Alexander Borbély

Why do we sleep? Why do we dream? Can we reduce our need for sleep? What can be done about insomnia? All these questions and more are answered in this enthralling exploration into the world of sleep.

April ISBN 0 582 00537 X Cased £10.95

ORGANIC CHEMISTRY FOR STUDENTS OF BIOLOGY AND MEDICINE

By Giles Taylor

A unique and comprehensive text, now revised and updated for this third edition, which provides the student of the life sciences with the basis for getting to grips with the chemical aspects of life processes.

May ISBN 0 582 44708 9 Paper £10.95

THE MICROBIOLOGY OF TERRESTRIAL ECOSYSTEMS

By Bryant Richards

Micro-organisms are often neglected in ecological textbooks. Here, Bryant Richards attempts to redress the balance by emphasising the role of soil organisms, especially fungi and bacteria, in maintaining productive ecosystems.

May ISBN 0 582 45022 5 Paper £17.95

MODERN SELECTIVE FUNGICIDES

Edited by Horst Lyr

A team of international specialists describe the properties and antifungal spectra of selective — and mainly systemic — fungicides, dealing with their practical applications in agriculture, horticulture and timber preservation. A definitive source for agronomists, foresters and biochemists.

May ISBN 0 582 00461 6 Cased £30.00 (est)



Available through your bookseller. For further details on these and related titles, write to Clive Weatherley, Longman House, Burnt Mill, Harlow, Essex, CM20 2JE, UK.

Thinking small and thinking big

John Postgate

Microcosmos: Four Billion Years of Microbial Evolution. By Lynn Margulis and Dorion Sagan. *Summit, New York/Allen & Unwin, London: 1986/1987. Pp.301. \$17.95, £12.95.*

DESPITE the publisher's claims for originality and novelty, the message of this well-intentioned book has been familiar to biologists for much of the present century. It is that microbes play, and have played, a dominant role in the sustenance and development of the biosphere. Thus they are a crucial component of a resilient network of interactions among living things which maintains the chemical metastability of the geosphere, on which terrestrial life depends. Microbes help to buffer the geosphere such that it changes only slowly, roughly coordinately with the evolution of the biosphere. Indeed, they were the evolutionary precursors of macrobes, which demonstrably retain anatomical residua of their microbial ancestors; and evolution itself, which involves accretory, convergent and divergent pathways, has not ceased simply because man, a self-conscious, self-examining species, has emerged.

If I judge the authors' intention correctly, they wished to develop these concepts in a popular, palatable form, illustrated and revised in the light of the knowledge acquired over the past few decades. For example, our own microbial residua are no longer seen to be just the leukocytes but include the ex-prokaryotic mitochondria; several other cell components probably stem from prokaryotes, and plant chloroplasts seem also to be of cyanobacterial origin; and there are hints that an endosymbiotic association between a eubacterium and an archaeobacterial host was the progenitor of eukaryotic cells. We know that gene transfer is rampant among prokaryotes such that the whole Kingdom has access to a common gene pool and is thus capable of large evolutionary jumps. There is evidence that populations on the periphery of the biosphere (dry, cold, hot or saline regions) positively require the existence of the mainstream biota and function not only as foci of divergent evolution but also as reservoirs for repopulation after ecological calamity. And so on.

How far have Margulis and Sagan succeeded in their laudable endeavour? In my view not at all, because they have allowed their own enthusiasms, together with some naively simplistic statements of the views of others, to lead them into a counter-productive mish-mash of impre-

cision and dramatic over-statement. For example, because we are derived from microbes, we *are* microbes: we are part of a "superorganism we call the microcosm" which is to become a "supercosm", ultimately to spread throughout the galaxy. The epithets of pulp science fiction abound: the "keen" perceptions of scientists have permitted "dazzling" advances; "spectacular" clues led to "astonishing" conquests with "staggering" implications for, amongst other things, how life coped with the "treacherous effect of light", or how "amazing" changes in oxygen tension left the biosphere "humming with the danger and thrill of free oxygen".

Oh my golly gosh! Such hyperbole might not matter if it did not lead to confusion. But metaphors and similes become hypotheses and theories, fact and speculation are deeply intertwined, anthropomorphisms and teleologies abound. Predictably the upshot is mysticism; try the following (p. 152):

As tiny parts of a huge biosphere whose essence is basically bacterial we — with other life forms — must add up to a sort of symbiotic brain which it is beyond our capacity to comprehend or truly represent.

I smelled a deity there. And I was right. She duly emerged several chapters later as that embarrassing Earth Goddess called Gaia; this is the "supercosm" the authors have been preparing us for. Unhappily I am allergic to mysticism, whether it comes openly with prayer book or obfuscated with science. Yet the blurb and preface writers have found all this revelatory and it has the blessing of the Prophet Lovelock. Impressive erudition and much stimulating thought went into the authors' grandiose vision; I wish they had solved the problem of putting it into words scientifically. □

John Postgate is Professor of Microbiology and Director of the Agricultural and Food Research Council Unit of Nitrogen Fixation, University of Sussex, Brighton BN1 9RQ, UK.

In the beginning

Sarah Hargreaves

The Creation of Life: Past, Future, Alien. By Andrew Scott. *Basil Blackwell: 1987. Pp.211. £14.95, \$19.95.*

APART from the continuing resistance from Creationists, the evidence for the evolution of life on Earth is well accepted. But how did life begin? In this book, Andrew Scott sets out not to attempt to give his own answer to the question, but to provide a sober appraisal of the competing ideas and the sparsity of the facts on which those ideas are based.

The book begins with introductory descriptions of background physics, chemistry and biology. Scott then tackles the "standard" theory, which was given impetus by the experiments of Stanley Miller in the early 1950s and which holds that life originated on Earth through the spontaneous reaction of organic compounds. This, according to Scott, is a "simplistic tale of the modern textbook and classroom" and he proceeds to identify its shortcomings. Although the basic constituents of living things could well have been abundant on the early Earth, there is no compelling indication that they were concentrated enough to generate nucleic acids and proteins; we have only sketchy evidence for the advent of replication, and that concerning the origin of the nucleic-acid-protein connection is even flimsier; finally, there is the question of the formation and evolution of the earliest cells, a process which no one has yet recreated in the laboratory.

What, then, are the alternatives? The following two chapters examine some

well-known possibilities. Greatest prominence is given to Graham Cairns-Smith's "low-tech" theory, according to which the earliest types of life were composed of inorganic, crystalline materials, able to replicate and evolve. These made possible the more familiar "high-tech" life forms based on nucleic acids, and were later taken over by them. Panspermia (the arrival of life on Earth from elsewhere), articulated by Arrhenius and later propounded by Francis Crick, also gets an airing, as does the view of Fred Hoyle and Chandra Wickramasinghe that interstellar space is full of life. Each of these ideas is discussed sympathetically and with consideration of how support for them might be obtained. Throughout, however, Scott sounds the proper notes of scepticism.

The discussion then turns to the development of complex organisms from the most basic of cells (with a brief and unsatisfactory foray into neurobiology) and the search for life elsewhere in the Universe. There is, too, a look to the future — to the research prospects on the origin of life, and to the possibilities opened up by genetic engineering, and by robot and computer technology.

This is a book of questions rather than answers, and the author has not attempted to provide a definitive guide to his subject matter. At times, particularly in the last two chapters, he passes over some tantalizing points all too quickly, but a helpful list of further reading is provided for those who wish to delve further. The book is lucidly written and is illustrated with clear, straightforward diagrams. It will serve very well as an introduction for the non-scientist who is intrigued by the puzzles surrounding the origins of life. □

Sarah Hargreaves is on the staff of Nature.

Pigafetta's paradox

John Brady

The Timing of Biological Clocks. By Arthur T. Winfree. *W. H. Freeman: 1987. Pp. 199. £15.95.**

THIS is a colourful book — in both senses of the word — embellished on every page with all the hues of the rainbow. But why such lavish illustration to analyse that insubstantial, grey element, time? The answer is that Winfree uses the spectrum as an analogy, so that each wavelength represents a different phase of the underlying oscillation that drives every biological rhythm, or at least describes its form.

The result is visually stunning. But whether or not it will help readers in understanding the subject matter will depend on their previous level of knowledge. And that raises the question of whom the book is intended for. The publisher is surely not aiming at the biological-clock *cognoscenti*, most of whom have followed Winfree's work over the past two decades (or have at least tried to). The informed man in the street, then? Hardly. He is likely to be lost before the end of Chapter 1, trying to fathom how the world's time-zones end in a 'singularity' at the South pole, yet walking round the pole has no effect on one's calendar, or why Pigafetta, Magellan's log-keeper, lost a day during his voyage round the world. I imagine that the book will appeal most to the mathematically sophisticated: readers of the *Journal of Theoretical Biology*, for example, or physical scientists with an interest in biology.

I once heard C. S. Pittendrigh — Winfree's doctoral supervisor at Princeton, and one of the three Grand Viziers of chronobiology — describe Winfree's work as 'baroque'. Presumably, he meant in the sense of 'exuberant and extravagant architecture'. The word came back to me repeatedly as I read the book, and grasped for the significance of converting a two-dimensional phase-response curve (which I understand) into a three-dimensional torus (which I don't), or when grappling with four-dimensional graphs of yeast metabolism (x , y , colour and numbered contour).

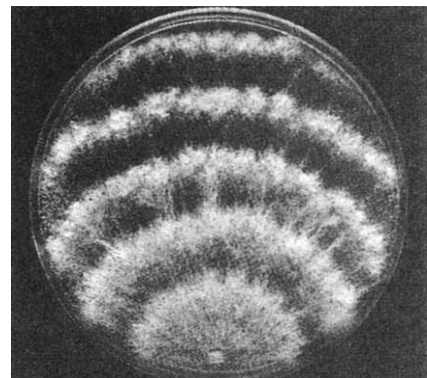
Certainly, what Winfree has done over the past 20 years has been to inject into the analysis of biological time-keeping a whole new approach to the subject. He has taken the single, linear dimension we were used to, and returned it to us stretched and folded into exuberant multi-dimensional arrays. While this has opened our eyes to the complexities of time, it is not clear how many of us have really

* In the United States published as a volume in the Scientific American Library.

benefited from the experience. There is also an element of the alternative, 'whimsical' sense of baroque in Winfree's delight in quoting from James Thurber and Lewis Carroll, which suggests, perhaps, an almost mischievous pleasure in juggling with time. Otherwise, any whimsy is more fairly identified as only the impression to be gained by the less-sophisticated reader; Winfree's analyses, though elegantly presented, are intricate, and the faint-hearted may choose the line of least resistance and pass by without comprehending.

Time is, indeed, a difficult subject. For scientists in general it commonly appears as the abscissa on their graphs, showing the progress or rates of reactions. For the chronobiologist, however, it often appears on both axes — especially as phase-response curves. Here it may no longer be linear, showing the amount by which the timing of some physiological process changes in response to a change in the timing of the environment. Time in two dimensions is hard to grasp, and the book is very much concerned with explaining the implications of such phase responses.

Winfree restricts his analysis to rhythms, and especially to the ubiquitous, 24-hour 'circadian' variety: seasonal day-length responses, hour-glass-type interval timing and time-sense of the kind necessary for navigation by long-distance migrants are not touched upon. The biological clocks of the title are therefore distinctly exclusive. Nevertheless, Winfree stands unique in the science of biological



Circadian advance — growth rings of *Neurospora crassa*, 22 hours apart.

time-keeping by visiting all fields, from population dynamics, through cells, to biochemistry. And he offers a grand synthesis that no one else could have attempted, although it must be said that his analysis is primarily concerned with the phenomenology of biological rhythms rather than their mechanisms.

For the biomathematician, this may be sufficient satisfaction in itself; for the physiologist, however, the former is justified only so far as it illuminates the latter. Winfree has provided some brilliant insights into how biochemical and cell systems may interact temporally, but it is not yet clear how far this approach will really accelerate our capture of a circadian clock; personally, I enjoyed his fascinating *Geometry of Biological Time* more. □

John Brady is Reader in Insect Behaviour in the Department of Pure and Applied Biology, Imperial College, London SW7 2AZ, UK.

Behind the curtain

Richard Mabey

The Natural History of the USSR. By Algirdas Knystautas. *Century Hutchinson: 1987. Pp. 224. £14.95.*

WHEN socialism was young, and still full of vision, nature was seen as part and parcel of the world that had to be liberated from capitalist oppression. Or, more precisely, helped towards its own liberation, to become Subject, not Object. The young Marx talked of the "resurrection" of nature. Bertolt Brecht waxed lyrical about the "otherness" of trees, the value of their having "something about them which cannot be put to use".

Then — for the Soviet Union at least — came war, Stalin, economic collapse and the kind of desperate, rushed ransackings of the natural world that climaxed at Chernobyl. Many in the West have assumed that more ghastly environmental catastrophes lie hidden in the recesses of that vast country, and that exploitation of nature is unavoidable in a self-styled materialist society.

The remarkable thing about Algirdas Knystautas's book is that it not only lifts the curtain and challenges many of these assumptions, but succeeds in recapturing something of the founding father's dream of a kindredness with nature. It is not a theoretical book, still less a political one. Yet even the statistics — those much parodied Soviet litanies — have the unfamiliarity and excitement of a survey of another planet. The Soviet Union has 3 million rivers, 2.8 million lakes, 100,000 species of plant and 800 species of bird. Many of these live in the 8 million square kilometres of forest that is popularly imagined to blanket most of the country. But the Soviet Union accounts for almost one-sixth of the Earth's land surfaces, and its habitats range from the south-western deserts, with their wild asses, gerbils and giant fennels, to the mysterious lush woodlands of Ussuriland in the Far East, where half of the country's bird species have been recorded. The Soviet Union is nothing less than the main reservoir of the ecosystems of the Northern Hemisphere, and it is no wonder that young, ecologically aware scientists should begin to demonstrate their concern about it.

Yet Knystautas is something of a

natural phenomenon himself. If it were not for the new mood of "reconstruction" he might be thought a mythical creature from a Soviet bestiary: a 30-year-old freelance, who writes elegant and witty English and takes his own accomplished photographs.

It is his pictures (and those by other Russian photographers) that take the book way beyond a bland catalogue. They are reflective, affectionate, respectful, often quite simply *awed*. Vast and entirely useless taiga swamps and unclaimed tundra roll into the distance. A Russian desman, a bizarre aquatic mole, squints into the camera. Exquisite wild onions and tulips are shown in their best finery. And up on shingly river banks in the mountains of Tien Shan stalks the extraordinary and never properly photographed ibisbill, half wader, half ice-pick. Knystautas's infectious excitement at new discoveries, new seasons, new feelings, fills his writing too. His enthralled account of finding a blue whistling thrush in the wild mountainscapes of Central Asia — a bird whose "trills are so piercing that they can be heard... even above the noise of the waterfall" — would have done credit to the best of the Romantics. This is a bird valued entirely for itself.

The same concerns are evident in the long chapter on conservation. Soviet programmes do, of course, have a strong utilitarian base. The laws on pollution control and species protection, and the stratified system of reserves and parks based on them, have identical aims to those in the West: "for the maintenance of genetic resources... protecting unique landscapes or habitats... the study of the processes in a completely natural community". I suppose that at an ungenerous pinch the conservation of woodlands near hospitals could be seen as utilitarian. But beyond this there is a more tender note, a delight in the variety of the natural world, and a concern about its fragility. And, yes, a real willingness on the part of the author to criticize his own country for disasters such as the pollution of Lake Baikal.

Knystautas is a maverick, but no dissident. His book was launched at the Soviet Embassy. Yet statements such as "human activity is now so intense that there is a real danger of disturbing the stability of the entire system, on which every one of us, rich and poor, depends" would have been unthinkable a few years ago. Of the many changes happening in the Soviet Union, this new perception is one of the most encouraging — for all Subjects of the Earth. □

Richard Mabey, 10 Cedar Road, Berkhamsted, Hertfordshire HP4 2LA, UK, served on the Nature Conservancy Council from 1982 to 1986. His most recent book, Gilbert White: A Biography of the Author of The Natural History of Selborne, was the winner of the 1986 Whitbread biography award.

Wild writings

David E. Allen

Penguin Nature Library.*

The Exploration of the Colorado River and Its Canyons. By John Wesley Powell. Introduced by Wallace Stegner. Pp. 407.

My First Summer in the Sierra. By John Muir. Introduced by Gretel Ehrlich. Pp. 264.

Cape Cod. By Henry David Thoreau. Introduced by Paul Theroux. Pp. 319.

Nature's Diary. By Mikhail Prishvin. Introduced by John Updike. Pp. 188.

"NATURE writing... is a very special... genre", claims Edward Hoagland, the general editor, in his foreword to these first four titles in the Penguin Nature Library. As broadly defined by him, it consists of writings which combine rhapsody with science — normally, he might have added, science that is conducted out of doors and for preference in wild and remote areas. An offshoot of romanticism, it rose to popularity in the second half of the last century and reached its high point in the inter-war years. Since then exponents have been rare, deterred, it seems, by the general switch in taste to greater matter-of-factness and by acute distrust of subjectivity instilled by a more intensive professionalism. It has always been a treacherous genre, exposed on the one hand to the Scylla of anthropomorphism and on the other to the Charybdis of the banal, and its reputation has been badly tainted by the many who have tried their hand at it without exercising the requisite control. Yet even so, as these four volumes illustrate, it *has* had its masters; and it was high time that a new generation of readers was introduced to what the best practitioners could achieve.

Except for the fact that they were all men who took delight in wilderness, the four authors who have been selected offer an appetizing set of contrasts. Thoreau and Prishvin have a focus that is primarily zoological, Muir was mainly a botanist and Powell an ethnographer and geologist. Thoreau writes of the sea-shore, Prishvin of forest and swamp, Muir of lush mountain scenery, Powell of arid river valleys. Prishvin's book is the journal of a year, Powell's (in the main) and Muir's are diaries of testing four-month trips, while Thoreau's is a collection of essays extending over a mere three weeks' comparatively gentle sauntering. All were engaged in exploring country that was new to them, but whereas the last three, all Americans, were on the move while they did so, Prishvin, a Russian — and the only trained

scientist of the four — was anchored to a single location.

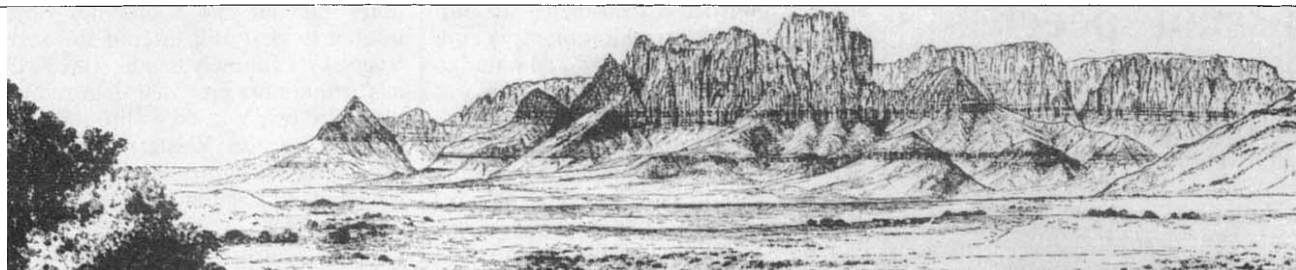
Powell's book is the closest to a traditional explorer's narrative and there is an element of suspense in his account which is lacking in the others. It deals for the most part with an expedition that he organized and led in 1869 down the chain of canyons through which the Green and upper Colorado Rivers run for over a thousand miles. Described by Wallace Stegner in his introduction as "the last great exploration within the continental United States", it opened up an area as big as Texas which had been penetrated thitherto only by Indians and a handful of Spanish adventurers and priests.

Ten men in four boats (purpose-built in Chicago) set forth in May, taking rations that they thought would last them as long as ten months; but mile after mile of falls and rapids gradually took their toll, and by August food was running short, all the instruments were ruined and three of the party were driven to breaking away — never to be seen again. Powell and his surviving companions eventually struggled home, but only after forfeiting most of the data that they had compiled. In essence the report which he subsequently produced for the Smithsonian, the book is marred by an over-lengthy scene-setting which takes up as much as four chapters. In compensation, however, it is handsomely decked out with the many original wood engravings which *Scribner's Magazine* so generously paid for, and there is abundant testimony both in these and in the text of the author's deep and abiding interest in the Indians and their customs.

John Muir's special passion was trees (he is the Muir commemorated in the famous redwood reserve north of San Francisco). By an odd coincidence the "forever memorable High Sierra excursion" of which he writes was undertaken in the very same summer as Powell's incomparably more arduous expedition. Muir had been hired to supervise the driving of a flock of sheep from the hot Californian plain up into summer pasture far above, a task which left him with ample leisure and allowed him to indulge to the full his favourite occupation of sketching. The word-pictures which he reconstructed so much later from his notes still retain the youthful excitement that must originally have gone into their making; and if "marvellous", "magnificent" and "grand" recur overmuch, there is a rapturousness in the descriptions (as Gretel Ehrlich notes) which confers on them a special quality.

As the introducer of *Cape Cod*, Paul Theroux was an especially happy choice, for not only was he raised in Massachusetts and has first-hand acquaintance with the area concerned, but he writes as a veteran contributor to travel literature himself.

*All four titles are available in the United States, price \$6.95 each. Only *Nature's Diary* and *Cape Cod* will be published in Britain (on 28 May, price £4.95).



Outstanding feature — "Towers of the Vermilion Cliffs", an engraving from John Wesley Powell's *The Exploration of the Colorado River*.

This little-known work by the famed author of *Walden*, he points out, is "unusual even by Thoreau's standards, because it is good-natured about the bleakest of subjects, it is only glancingly about the Cape (its real subject is the sea), and contains more reverie than actual experience". Essentially it is a book about beachcombing — but beachcombing elevated almost into a spiritual commitment. Like W.H. Hudson, that self-styled "traveller in little things" with whom he has sometimes been compared, Thoreau specialized in the sort of minor observations and incidents which most other authors consider beneath their regard, allowing his narrative to wander about here and there, as seemingly aimless as his feet. Between them the two constitute nature writing's picaresque wing.

Though stereotyped as a loner, Thoreau gives evidence enough in this book of at least being interested in the people that he happened to encounter. What is more, on two out of the three short excursions (made in 1849–1855) on which it is based he had a friend as his companion. Even worse for the image, the book was a product of what would appear to have been a calculated bit of worldliness: it originally consisted of ten separate articles, written at a time when he was supporting himself by lecturing and journalism. Eleven years after *Walden* and three years after his comparatively early death, the unity of the pieces was recognized and they were successfully disinterred. Quite apart from their literary attractions, they possess a certain interest for the historian as well. For they were written at just about the time that Philip Henry Gosse was attracting an unexpectedly large following with his entrancingly illustrated guides to the life of the seashore on the other side of the Atlantic. It seems more than possible that Thoreau's publisher, aware of this, decided to lure him away from his hitherto land-locked existence and try to produce in him an American answer to Gosse. "Having come so fresh to the sea", Thoreau is frank enough to admit, "I have got but little salted". Did he realize, one wonders, that he risked disappointing posterity by showing himself as opportunistic?

Unlike the other three books, Prishvin's *Nature's Diary* is a work of the twentieth century. Unlike them, too, it is a trans-

lation and it comes from a different continent. As its introducer, John Updike suffers from the disadvantage of not knowing the area described and is reduced to reminding us of the scenery at second hand, through the eyes of such as Tolstoy and Turgenev. Much-travelled, widely read, a one-time student in Leipzig, Prishvin had an extra coating of sophistication to shed before he was able to approach nature with the necessary directness of response. His book is not a continuous narrative, but consists of unrelated journal extracts which inevit-

ably give it a rather disjointed character. Its affinity is rather with *belles lettres* and with the selfconscious nature essayists than with the higher travel-writing of Muir and Powell. Devotees of Thoreau may nevertheless find it an interesting comparison — though they are unlikely to approve of the author's egg-collecting and his frequent resort to the gun. □

David E. Allen, *Lesney Cottage, Middle Road, Winchester, Hampshire SO22 5EJ, UK*, is a past President of the Society for the History of Natural History. His most recent book is *The Botanists* (St Paul's Bibliographies, 1986).

Doom again

Kenneth Mellanby

Earth. By Anne H. Ehrlich and Paul R. Ehrlich. *Thames Methuen, London/Franklin Watts, New York: 1987. Pp.258. £14.95, \$19.95.*

AN UNKIND reporter of a scientific conference is often tempted to say "Professor Blank gave his lecture". Similarly I am tempted to say "Professor Ehrlich has published his book". Over the past 20 years, Paul Ehrlich, sometimes alone, sometimes in collaboration with his wife, Anne Ehrlich, or with one of his other colleagues, has written a series of books all of which contain essentially the same information. Only a year ago he produced *The Machinery of Nature*. In 1966 his first major contribution was *The Population Bomb*. In between we have had *Population, Resources, Environment* in 1970, *Ecoscience* in 1977 and a great many articles in many different magazines.

It would be wrong to say that these books are all only slight modifications of the same text. Like a skilled chef, Ehrlich has used the same ingredients to produce a considerable variety of dishes. For one thing, the books vary considerably in size and in the detail in which the information is presented. They are intended to appeal to different audiences, from professional scientists to the general public. *Earth* is perhaps the most popular of all, and is aimed at the audience expected to watch a television series shown in Britain earlier this year. Like all his books, it is well written. It is rather more lavishly illustrated

than most of the earlier works. Had they not been published, this new one would for the most part have been warmly welcomed.

The information presented in so many ways in Ehrlich's books consists first of an introduction to global ecology, well and clearly expressed. We are then given a picture of the world today, with particular stress on the terrible things that mankind is doing to the environment. Finally we have a forecast of the future. My main criticism is that when describing the world today, the most gloomy picture is always given, and any report of damage, however tentative, is stated as fact.

There has, however, been some change over the years. In 1969, in an article entitled "Eco-catastrophe", Ehrlich forecast the death of the oceans, and "almost instant starvation" in Japan and China well before 1980. He makes no references to this and other dreadful but unfulfilled prophesies in his more recent works. Although still a prophet of doom, the picture is a little less pessimistic than that presented 20 years ago. Overpopulation by man is still seen to be the greatest danger, but the possibility of population control within the world's carrying capacity is no longer considered impossible.

Professor Ehrlich is a highly respected ecologist, and he has made many valuable contributions to science. It is unfortunate that he is unable to exercise the rigour he clearly uses in the laboratory when he becomes involved in world affairs. □

Kenneth Mellanby, 38 Warkworth Street, Cambridge CB1 1EG, UK, was formerly Director of Monks Wood Experimental Station, Huntingdon. Among other books he is author of *Farming and Wildlife* (Collins, 1981).

Russian speech

John C. Marshall

Thought and Language. By Lev Vygotsky. Revised and edited by Alex Kozulin. MIT Press: 1986. Pp. 287. Hbk \$27.50, £24.75; pbk \$9.95, £8.95.

WHEN Vladimir Mayakovsky killed himself in 1930, the great Russian linguist Roman Jakobson commemorated his achievements in an impassioned essay "On a Generation That Squandered Its Poets". And not only its poets.

Between the Fourth Duma, summoned in 1912 by Tsar Nicholas II, and the imposition of Stalin's tyranny in the late 1920s, the creative potential of the Russian peoples flowered as never before (or since). Despite war and revolution, censorship and economic chaos, the intelligentsia somehow managed to keep in close contact with the *avant-garde* of Western Europe whilst making signal contributions to all the arts, to linguistics and to psychology. The charismatic leader of Soviet psychology in this generation that had the misfortune to live in interesting times was Lev Vygotsky.

Born near Minsk in 1896 (the same year as Jakobson), Vygotsky graduated from the local Jewish gymnasium with honours

and was awarded a gold medal for outstanding all-round achievement. Yet it was by sheer luck that he was admitted to Moscow University; the Jewish quota was 3 per cent — to be chosen by lottery. The dice fell in Vygotsky's favour. He intended to study philosophy (Hegel and Spinoza, in particular), but at his parents' prompting ("From philosophy you can earn a living?") entered first the Medical School and shortly thereafter transferred to the Law School. His law studies did not, however, prevent him from participating fully in the artistic life of Moscow, nor from enrolling at a private university, Shaniavsky, from which he graduated in history and philosophy.

After his student years in Moscow (1913–1917), Vygotsky moved to Gomel (1918–1924), where he first taught school children and then lectured at the local teachers' college. It was here that his interest in developmental psychology and in the education of the physically and mentally handicapped began to dominate his work. An earlier fascination with literary creation and criticism persisted, and Vygotsky concurrently finished a dissertation, *The Psychology of Art*, for which he was awarded a PhD from the Moscow Institute of Psychology.

Professional recognition came in 1924, after a lecture he delivered to the Second Psychoneurological Congress in Leningrad. This talk — "The Methodology of Reflexological and Psychological Studies" — attacked the official Soviet doctrine that 'consciousness' was merely "an idealist superstition", as Kozulin writes, suitable at best for the "nonscientific psychology" in which one's grandmother is likely to believe. The lecture made a profound impression upon Alexander Luria, whose later studies of the cognitive effects of traumatic brain damage sustained by Russian soldiers in the Second World War are one of the cornerstones of modern neuropsychology (and in all probability saved him from the Gulag). Luria was instrumental in obtaining a position at the Moscow Institute for Vygotsky, and always regarded his own work as a continuation of Vygotsky's.

For the remaining decade of his life, Vygotsky worked and wrote like a demon, despite severe illness and the odium of both the psychological and political establishment. Although he did publish during his lifetime, at least 80 manuscripts remained unpublished and only began to surface in the Soviet Union after 1956.

Thought and Language, Vygotsky's last and best-known work, appeared in the year of his death (1934), but was suppressed, along with all his other writings, two years later. This new English edition is required reading for all psychologists, East and West. The translation has been thoroughly overhauled, and extensive

notes fill out the sometimes obscure references that still littered the text on Vygotsky's untimely death. Alex Kozulin has furthermore provided an introductory essay that sets Vygotsky's life and work in context for the Western scholar who expects his or her own work to provoke no government reaction more extreme than benign neglect.

What claims, then, does *Thought and Language* make, and why should they have been regarded as subversive? Vygotsky attacks the work of Jean Piaget, and argues that the young child is not as 'illogical' as Piaget makes out, provided that he or she is familiar with the situations, words and things that the investigator is manipulating. Different societies and social milieux will provide differential access to knowledge and experience. Hence, as Vygotsky writes, "One may imagine how substantial must be the differences in data collected in Genevan and Soviet kindergartens". Should Vygotsky have said that the Soviet data were 'better' than the Swiss, rather than merely different?

Vygotsky then criticizes William Stern's account of the child's first words. For Stern, the single word "Mama" can be translated as "Mama, come here" or "Mama, put me in the chair". No, says Vygotsky, there is no evidence for such intellectual underpinnings. Thought, affect and language develop independently. Only later do they converge in a 'symbolic' system that enables the child to express internally (and overtly) thoughts about his situation. This convergence eventually allows the development of thoughts that are not directly controlled by the child's environment. One begins to see why the authorities were displeased with the new psychology.

Equally objectionable was Vygotsky's interest in the poet's search for the exact word to express publicly the inchoate thoughts and feelings of an individual citizen. The last chapter of *Thought and Language* is prefaced by lines from Vygotsky's fellow 'cosmopolitan', Osip Mandelstam:

The word I forgot
Which once I wished to say
And voiceless thought
Returns to shadows' chamber.

When Vygotsky was dying of tuberculosis in a Moscow hospital, the secret police were already knocking on the door of Mandelstam's apartment. Yet it was precisely the poets, painters and musicians who nourished the people's capacity to resist in the darker days that were yet to come. It was Anna Akhmatova who wrote, during the siege of Leningrad:

And we will preserve you, Russian speech,
The great Russian word. □

John C. Marshall is in the Neuropsychology Unit, part of the Neuroscience Group at the Radcliffe Infirmary, Oxford OX2 6HE, UK.



Primate Societies

Edited by BARBARA B. SMUTS,
DOROTHY L. CHENEY,
ROBERT M. SEYFARTH,
RICHARD W. WRANGHAM &
THOMAS T. STRUHSACKER

With contributions from 46 experts, spanning the disciplines of behavioural biology, ecology, anthropology and psychology, *Primate Societies* is the most systematic, detailed description and analysis available of social behaviour across the entire order of nonhuman primates.

£48.00 Cloth 608pp illus.
0-226-76715-9
£19.95 Paper 0-226-76716-7
Publication: 3 July

THE UNIVERSITY OF CHICAGO PRESS

126 Buckingham Palace Road London SW1W 9SD

Reader Service No.42

Totally out of the question

Stuart Sutherland

No Way: The Nature of the Impossible. Edited by Philip J. Davis and David Park. W.H. Freeman: 1987. Pp.325. \$17.95, £17.95.

RUDYARD Kipling wrote:

I'm the Prophet of the Utterly Absurd,
Of the Patently Impossible and Vain.

The banjo's boast could be the theme song of creative scientists, who frequently make what seems impossible one day possible the next. Once upon a time it seemed impossible that the Universe was not eternal, that the Earth went round the Sun, that there was a finite velocity beyond which nothing could travel and that there could be a geometry that did not make the assumption that parallel lines never meet. Had there been anyone around to survey the Big Bang, he would surely have thought it quite impossible that one of its distant products would be mankind.

Professors Davis and Park have gathered together contributed chapters on 18 different subjects, ranging from chemistry to music and from biology to law. Each chapter sets out some of the impossibilities within the discipline it covers. The contributors for the most part write aimably and clearly. They often provide fascinating glimpses of their own subjects, though they rarely illuminate the nature of the impossible.

It might have been helpful if the editors had distinguished more carefully between different kinds of impossibility. Within science, what is thought to be impossible becomes possible if the theoretical framework is changed. Physics successfully absorbed the fact that matter can be converted into energy. But — a point hardly noticed in the book — there are different degrees of impossibility. Suppose, for example, foreseeing the future through clairvoyance were found to be possible. The business of bookmakers and of allied professions such as stockbroking would be severely damaged, but of more importance physics would find it impossible to absorb the finding for it runs counter to one of its most far-reaching assumptions. The finding could not just be explained by developing a broader theory in the way newtonian physics falls out as a special case of relativity theory; it would either require a complete rethinking of the whole of physics or — more likely — would have to be quietly ignored.

Apart from the impossibility of breaking the existing rules of a science, there are technological impossibilities which depend either on theoretical limits or on

the limitations of practical techniques. For theoretical reasons, based on heat loss, a petrol engine cannot convert more than about two thirds of the energy in the petrol into mechanical energy to propel a car, but for practical reasons the efficiency of existing engines is very much lower than this. Similarly the switching times of gates on a computer are limited by thermal fluctuations which create errors, but the present rate at which operations can be performed (about 10^7 per second) is well below the theoretically possible maximum. Yet even at a practical level the impossible often becomes possible. Although these topics are not mentioned in the book, it is not so long ago since heart transplants and blood dialysis would have been scoffed at. Surgeons are already at work to provide an artificial womb so that men can give birth, but even if the obscenity of some medicine makes it possible for a man to bear a baby one can feel fairly certain that babies will never bear men.

The impossibility of falsifying propositions that are true by definition (for example " $2 + 2 = 4$ ", or "nothing can be red and blue all over") is less interesting, except that even here the framework may change — as in the case of parallel lines never meeting — and life goes on.

One of the most interesting chapters examines how the law deals with impossible stipulations under a contract. In 1705 someone, who was clearly no mathematician, undertook to supply two grains of rye each Monday and to double the amount supplied on each successive

Monday for a year. He tried to escape the contract on the grounds that it was impossible to fulfil, but the judges took a different view. One announced "where a man will for a valuable consideration undertake to do an impossible thing, though it cannot be performed, yet he shall answer damages". The other more tersely pronounced "the defendant ought to pay something for his folly". Although the issue of impossible contracts has not been fully resolved, the law tends nowadays to take the sensible view that if someone knowingly commits himself to performing the impossible, he should be responsible for his promise, but if he has no reason to believe it to be impossible he may be absolved.

There is a devastating chapter on education that rightly notes that it is not a science and that not a single finding or theory in psychology is of any use to the practising educator. The chapter on psychology concentrates too heavily on a well worn theme — the impossibility of being a good parent. It might have been more interesting to discuss wider issues, such as the impossibility of finding happiness from within or the impossibility of deliberately changing oneself. But no book can discuss all impossibilities, which range from "you can't get an ought from an is" to "you can't get a man with a gun". Although *No Way* is not impossible to put down, it is impossible to resist dipping into it. □

Stuart Sutherland is Director of the Centre for Research in Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.

Against ignorance

Patricia Smith Churchland

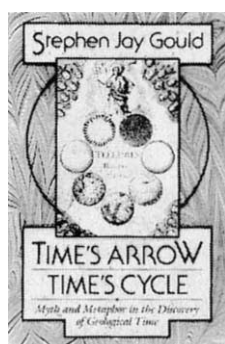
Philosophy and the Brain. By J. Z. Young. Oxford University Press: 1987. Pp.233. £12.95, \$22.50.

NEUROSCIENCE has progressed to the point where it has begun to impinge seriously and systematically upon our understanding of human nature. Although nothing like a full-scale theory has emerged, there are many entry-points where we can begin to make neurobiological sense of psychological states and processes: what it is to see, be awake or dreaming, feel pleasure and pain, and so forth.

Many philosophers have started to take on board the implications of recent developments in neuroscience, but there remains a diehard core for whom neuroscientific discoveries are irrelevant. Favouring the *a priori* approach to questions about what we are and how we work, they use the label 'scientism' for the view that neuroscience can provide enlightenment, even as — perhaps especially

as — it changes our intuitive conceptions of ourselves. By hitching the pejorative 'ism' to the end of an otherwise respectable word, they convey the suggestion that science is encroaching on the territory of the humanities, where, apparently, it has no legitimate business. To such people the very idea that behaviour might have a neurobiological explanation seems demeaning to human dignity and self-esteem.

In *Philosophy and the Brain*, J. Z. Young tries to convince the sceptics that if they allowed themselves to know what research has revealed in the neurosciences, they would recognize that the *a priori* approach is blinkered and stifling. They would see, too, that neurobiological explanations of human character and behaviour enhance rather than detract from the dignity of man, because they give us insight and understanding based on fact. Perplexed and dismayed by the policy of ignorance, Young has tried to assemble a sort of neuroscience-without-tears. By keeping the book short and minimizing technical language, he has tried to make it easy for philosophers to see why they should care about how the brain works. ►



Time's Arrow, Time's Cycle

Myth and Metaphor in the Discovery of Geological Time

STEPHEN JAY GOULD

Gould's subject is geology's signal contribution to human thought – the discovery of "deep time", the vastness of earth's history.

£13.95 Cloth 240pp illus. 0-674-89198-8
1 May

Metabolic Arrest and the Control of Biological Time

PETER W. HOCHACHKA & MICHAEL GUPPY

This pioneering book examines the ability of some animals to regulate their own metabolisms.

£21.95 Cloth 256pp illus. 0-674-56976-8

Walter B. Cannon

The Life and Times of a Young Scientist

SAUL BENISON, A. CLIFFORD BARGER & ELIN L. WOLFE

The first biography of one of the 20th century's most distinguished crusaders of medical science.

Belknap £23.95 Cloth 528pp illus.
0-674-94580-8 July

Revolution in Science

I. BERNARD COHEN

'Monumental scholarship... In sheer variety of riches this book has few parallels.' *Nature*

£7.95 Paper 736pp illus. 0-674-76778-0

Niels Bohr: A Centenary Volume

Edited by A. P. FRENCH & P. J. KENNEDY

'If you have any interest at all in 20th-century science, then it is essential reading.' *New Scientist*

£11.95 Paper 424pp illus. 0-674-62416-5

The Dialectical Biologist

RICHARD LEVINS & RICHARD LEWONTIN

'...an extremely readable and very provocative book.' *New Scientist*

£7.25 Paper 320pp illus. 0-674-20283-X

The X-Ray Universe

WALLACE TUCKER & RICCARDO GIACCONI

'Scintillating... the first readable and comprehensive account of the X-ray universe.' *Times Higher Education Supplement*

£6.25 Paper 220pp illus. 0-674-96286-9

HARVARD

UNIVERSITY PRESS

126 Buckingham Palace Road London SW1W 9SD

Reader Service No.28

They should care, if only because neuroscience is directly relevant to Cartesian dualism (the claim that the mind is a separate substance, distinct from the physical brain, in which inhere all our mental states and processes). Young's general approach here is straightforward: the empirical data make dualism highly implausible. Once we see the systematic connections and dependencies between states described at the psychological level and states described at the neurobiological level, dualism looks about as tenable as vitalism. Young follows this up with the claim that dualistic implications are sometimes built into our psychological concepts, for example 'the self' and 'free will', and when they are, he avers, we ought to divest ourselves of the misleading implications.

Having argued for the basic thesis that mental states are states of the physical brain, Young proceeds to draw upon research in the neurosciences to display vignettes of what has been learned about the specific neurobiological basis of mental states and processes. Echoing themes from his earlier work, *Programs of the Brain* (Oxford University Press, 1978), he takes as an organizing principle the idea that nervous systems are fundamentally geared to enable the organism to act: principally, to feed, flee, fight and reproduce. The brain, as he puts it, has 'programs' for behaviour, and even something as ostensibly passive as vision should be seen in the context of action, as the search for information. Young takes this as the key to understanding knowledge in general: "my thesis is that human knowledge can be considered as a special development of the process of gathering information for life that is essential in all organisms" (p.79).

From this perspective, Young touches on a variety of results from physiology (such as the topographic and columnar organization of sensory cortex) and from psychology (visual hyperacuity and the ability of infants to respond differentially to all known phonemes). Along with many linguists, he takes the research on infant phoneme recognition as support for nativism for language. One problem with this conclusion is the finding that the chinchilla, of all unlikely animals, can, like the human infant, respond differentially to all the same phonemes. Consequently, whatever the innate substrate for this capacity, it does not appear to be unique to language users.

The book readily acknowledges that neuroscientists possess only pieces of the puzzle of how brains represent and process information. What we need to know is how networks of neurons interact to produce certain effects, but, as things stand, we don't have the physiological techniques for investigating properties at the level of the circuit or the system.

Modelling is, therefore, an important means of developing hypotheses at these higher levels. Young discusses Marr's approach to modelling, though he says very little about the connectionist approaches that have superseded Marr. What he does say is curiously unenthusiastic, perhaps because the new learning algorithms (Boltzmann machine and back-propagation) had not been formulated when he was writing the book.

In rounding out the general 'physicalist' picture, Young anticipates the time when we will have neurobiological explanations for sensory and cognitive processes, and also for human ethical behaviour and aesthetic preferences. Unafraid to grasp the nettle, he believes that as neuroscience proceeds and we understand more accurately the springs and principles of human behaviour, we may improve our ethical attitudes towards responsibility and blame. His argument rests on the thesis that insofar as cultural and social factors affect ethical behaviour, they must be represented in the brain, and such representation will eventually be within the compass of neuroscience. This is not so much an argument as a reaffirmation of 'physicalism', and Young tends to underestimate the difficulty of arriving at neuroscientific explanations of moral behaviour.

Although the aim of the book is admirable, there are weaknesses in execution, principally because Young skims very fast over both the philosophy and the neuroscience. On the philosophical side, he probably does not know the opposition as well as he might and regularly leaves himself open to accusations of naivety. On the neuroscientific side, most of the discussions are insufficiently developed to convey a genuine understanding of the research concerned. A further drawback to the book for the neuroscientifically innocent is that there are no illustrations — how can even the most diligent philosopher master the significance of hippocampal architecture or the widespread projection system of the locus coeruleus if he cannot see what it looks like?

Nevertheless, *Philosophy and the Brain* is the work of a man who has thought long and deeply about the neurobiological basis for psychological processes, and it gives an important view of the need to integrate philosophy and neuroscience. Young's insights into matters where empirical results bear upon philosophical opinion should at least move philosophers to wonder whether devotion to the *a priori* strategy has not been profoundly misguided. □

Patricia Smith Churchland is a Professor in the Department of Philosophy, University of California at San Diego, La Jolla, California 92093, USA, and author of *Neurophilosophy: Toward a Unified Science of the Mind-Brain* (MIT Press, 1986).

Trials of the gluttons for punishment

Harold McGee

Never Satisfied: A Cultural History of Diets, Fantasies and Fat. By Hillel Schwartz.
Free Press, New York/Collier Macmillan, London:1987. Pp.468. \$19.95, £14.95.

FOR Lord Byron it was hard biscuits, soda water *sans* hock and vinegar-drenched potatoes. For Louis Armstrong it was orange juice, antacids and laxatives. For Henry James it was "fletcherizing": slowly chewing the flavour out of each mouthful, which in the case of a shallot could take seven minutes. "Grapple it to your soul with hoops of steel", he wrote a friend (though five years later he would complain of a persistent "stomachic crisis: the condition of more and more sickishly *loathing* food"). Even today a distant echo of "The Great Masticator" Horace Fletcher can be heard issuing from an electronic scale "with a General Instrument voice synthesizer, 400k bits of ROM memory and a Hitachi microprocessor, saying 'Congratulations, I am pleased. You are making progress. Today practice taking more time to chew each mouthful'."

Hillel Schwartz's entertaining book is replete with anecdotes and curious, sometimes appalling facts from the history, mainly American, of weight-reducing regimens. The first systematic American scale-watchers, around 1835, were followers of the Reverend Sylvester Graham who felt obliged to prove that they would not lose weight on a temperate and vegetarian diet. Only at the turn of the century did slimness become desirable, and remedies formerly deployed against dyspepsia were turned against newly unwanted fat.

The quantification of nutrition was pioneered at about the same time by the chemist W. O. Atwater, who analysed and published the protein, carbohydrate and fat contents of common American foods. Beginning in 1904, menus at the sanatorium run by J. H. Kellogg, the inventor of corn flakes, gave calorie counts for each dish, and in 1906 the Yale economist and amateur gymnast Irving Fisher suggested replacing the ounce with 100 large (kilo-) calories as the standard serving measure. By the time of the First World War calorie-counting had been incorporated into several popular diet books. Kitchen scales were introduced in the 1920s, a couple of decades after bath scales.

Because it requires no active intervention of will power, an attractive alternative to caloric restriction has been pharmacological tinkering with metabolism. Iodine in the form of the seaweed *Fucus* was discovered as a reducing agent in 1859, and thyroid extract in 1883; both were incorporated into obesity cures early in the next century. Dinitrophenol, which uncouples oxidative phosphorylation and

dissipates energy stores as heat, was tremendously popular in the 1930s despite having been recognized as a toxin in munitions factories; it was banned in 1938. Ready to take its place was benzedrine, originally marketed as an asthma remedy, and the first of the amphetamines to be used for appetite suppression, even in



Chewing guru — Horace Fletcher, "The Great Masticator", whose white clothing reflected an equal concern with intestinal asepis.

two-year-olds. The digestive equivalent of contraception was proposed in 1977 by two physicians who discovered that perfluorooctyl bromide, a fluorocarbon used as a radiological contrast material, can coat the intestinal mucosa and limit nutrient absorption.

Never Satisfied is a lively and valuable contribution to the history of American foodways. Schwartz has gathered a massive amount of material from the popular, medical and historical literatures, and his book is exhaustively documented; a reference is given for nearly every sentence. He reconstructs what he calls the "shared fictions of the body" that have prevailed over the past 150 years, and demonstrates the contributions of chemists and physiologists, economists and insurance companies. He brings order to the bewildering variety of diets, sorting them by rationale and pointing out frequent recurrences, the "blue-smoke and trap-door entrances and exits".

Schwartz also aims to explain "the cultural fit between shared fictions of the body and the reducing methods of the era", but generally I find these correlations hard to swallow. The problem is that many methods co-existed in each era. For example, Schwartz describes fasting, slow-chewing, calorie-counting and thyroid pills as four contemporaneous methods that "dealt with gluttony and

fatness as if they were problems of flow" in a body characterized by "dynamic balance". Looked at less abstractly, they seem most remarkable for their differences. One could eat next to nothing or moderately or well; one exercised one's self-control, jaw muscles, arithmetic or thyroid gland; food was a forbidden temptation, a raw material in need of processing, a set of numbers or fuel to be burned. As Schwartz candidly admits at one point, "The truth was a mess. People set out to lose weight any way they could, regardless of compulsive historians, and they mixed everything up".

Schwartz's principal thesis is that slimming

is the modern expression of an industrial society confused by its own desires and therefore never satisfied. On the one hand we seem to want more of everything; on the other hand we are suspicious of surplus. Increasingly perplexed or intimidated by abundance, Americans have taken the protocols of slimming as the protocols for social and spiritual renewal.

Surprisingly little evidence is mustered in support of this proposition. Schwartz does locate the turn towards slimness during a period of economic overproduction and underconsumption, but he gives just as much credit to the "kinaesthetic" examples set by airplane pilots, efficiency experts and Isadora Duncan:

The modern dancers in the long run had probably the greatest effect, as their principles of movement became part of the cinematic code and established the standards for the motor education of most young children.

This important assertion also stands unsupported.

A more complete analysis of the appeal of slimming would place such regimens in the broader context of American health and diet fads. Lacking entrenched national traditions in belief or in social structure, the New World seems to have encouraged an optimistic sense that renewal, personal as well as "social and spiritual", was possible, but also an anxiety about seeing such renewal come to pass. In religion, this outlook fostered a great variety of sects, including several that proclaimed the imminent return of Christ. In dietary matters, abetted by the lack of a strong, unifying culinary tradition, it fostered a susceptibility to regimens that promised perfect health and long life. Such regimens flourished long before the triumph of slimness, and surely helped prepare the way.

A comparison of dietary trends in the New and Old Worlds would have brought these and other factors into relief. Schwartz quotes the joke, originally aimed at the English, of an early French visitor: "What a country! Fifty religions and only one sauce — melted butter!". By contrast, Catholic, generously sauced France resisted dietary change. A 1909 French textbook of domestic science, paraphrased

by the historian Theodore Zeldin

pointed out the danger of over-rich sauces, but insisted nevertheless that everybody should know how to prepare them and that no dinner was complete without them; the serving of appetising food, its artistic preparation and display were held up as one of the most important duties of every housewife.

Stephen Mennell, a sociologist, has pointed out that English cookery columns for the lower-middle classes "emphasized the need to eat fat and heavy food for body-building", and French women's magazines began calorie-counting only in the 1950s. "Abundance" or affluence is clearly part of the story, but not the whole story.

Schwartz's treatment of the role of science and medicine in dietary faddism is disappointingly sketchy. As he demonstrates, scientific findings have repeatedly provided the raw material for new fads; and many of the more egregious regimens have been and continue to be concocted by medical doctors. One would like to know how it has been that technical advances have found their way into the popular imagination. Who has done the bending to fit them to the desire for a panacea? Why have "health food" stores come to stock capsules of soy phosphatides, chelated copper and L-phenylalanine alongside their "organic" lettuce? Unfortunately, Schwartz finishes by lumping physiologists together with evangelical

Christian dieters (*More of Jesus and Less of Me*) and "behaviourists" (*How to Be a Winner at the Weight Loss Game*), all of them having "drawn obesity away from the perturbing social and economic issues of consumption". The neurophysiologists "hoped" to do this "by drawing appetite deep inside the brain". This is a novel species of scientific motivation.

Never Satisfied does not conclude with a straightforward account of current views on body weight and health, or on the amenability of weight to dietary manipulation. Perhaps Schwartz feels that today's answers are still provisional and will be out of fashion tomorrow. Instead, the last chapter opens with an all-out diatribe against slimming, complete with such remarks as "few militarists, murderers, or rapists are fat". This turns out to be a calculated exercise in imagining a future manifesto of the Fat Power movement, and the last pages note some current signs of discontent with the culture of thinness. Neither the author nor his reader finishes fully satisfied. Still, Schwartz is to be thanked for laying on quite a spread as he vividly reminds us that "Our bodies and our foods are as much social constructs as they are proteins, carbohydrates, fats". □

Harold McGee, 838 La Jennifer Way, Palo Alto, California 94306, USA, is author of *On Food and Cooking: The Science and Lore of the Kitchen*, published in the United States by Charles Scribner's Sons and in Britain by Allen & Unwin.

Fearful symmetry

David R. Rosseinsky

Symmetry Through the Eyes of a Chemist. By István Hargittai and Magdolna Hargittai. VCH:1986. Pp.458. DM156, \$90.30.

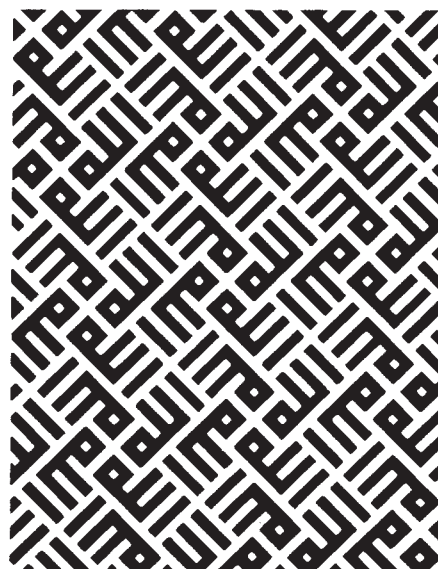
THE COMMON problem in early winter, of finding two left- or right-handed gloves, would be instantly resolved by dropping the arbitrariness in mitt design of differentiating interiors from exteriors. The inside-out remedy here is perhaps the only everyday inversion not drawn on by the authors in their lead-in to symmetry in science.

While Blake's Tiger! Tiger! is also missing, his God as Geometer provides a typical illustration, and Escher's obsessive collocations are yet again called upon. But many of the everyday illustrations are photographs of carefully observed, basically commonplace scenes picked out for some symmetric facet: the snowflake gets 10 pages. These examples provide a beguilingly attractive introduction, especially to the minority of readers deficient in three-dimensional perception (personal research shows us even to be perhaps a silent majority). Such a utilitarian assessment undervalues the aesthetic pleasure

to be derived from the arrays of examples of symmetry in art, sculpture and nature. Progress towards the general introduction of symmetry labels and symmetry operations follows naturally, together with their application to chemical systems.

Conical (*conical?*) periodic tables have however already featured in the introduction, and 1,2-dibromo-1,2-dichloroethane slips in early in Chapter 2. Then acetanilide and *p*-chloroacetanilide, with contrasting molecular arrangements in their centrosymmetric and noncentrosymmetric rhombic crystals, and chiral systems in general exemplified by 4-(bromomethyl)-6-(mercaptopomethyl)[2.2] metacyclopentane together with larger combined structures ease the reader towards quite elaborate chemical systems. Point groups are exemplified both by patterns and by molecules of greater or lesser complexity. The invaluable framework embodied in the valence shell electron pair repulsion model is discussed in some detail.

Group theory employing matrices is included, complete with a brief introduction to matrix manipulation, irreducible representations and character tables. Symmetry rules in vibrations precede electronic structure and symmetry in wave mechanics, and consideration of reaction mechanisms; symmetry rules for the reac-



Period return — drawing, by the crystallographer Khudu Mamedov, of the decoration on a mausoleum in Badra, Azerbaïdzhane, in which the basic motif is the word Allah.

tion coordinate are summarized and the requirements of orbital symmetry, following the views of Fukui and of Woodward and Hoffman, are exemplified in numbers of simple examples. Space-group symmetries in infinite systems are addressed largely by reference to non-chemical examples, and crystals and crystal groups follow, to give all the applications of symmetry in chemistry conceivable (to me).

This pleasant evocation of symmetry applications is accomplished with continued reference to up-to-date chemistry. Thus I find a recent and not entirely welcome reference to the probable non-linearity of gaseous transition metal dihalides: gone the simple coulomb view of bonding here, together with some cherished but now invalidated preconceptions.

The only non-chemical importations which in my view do not succeed are the passages from fellow-Hungarian writers. In these, verbal dualities and ambiguities of interpretation are explored, but out of context and in what seem contrived circumstances.

The margins take one-third of each page and attractively accommodate smaller diagrams, or the legends to larger textual ones, or otherwise simply present an aspect of spaciousness. Such extravagant counter-austerity (not of course new) will undoubtedly spread.

Almost education by stealth, certainly education by aesthetic appeal, the whole exercise is epitomized in the quoted words of Coulson: "Symmetry is important, but it is not everything.... It is when symmetry interprets facts that it serves its purpose because it links our study of chemistry with another world, [that] of the human spirit". □

David R. Rosseinsky is a Reader in the Department of Chemistry, University of Exeter, Exeter EX4 4QQ, UK.

Shenanigans in the market place

Anthony Sampson

Big Blue: IBM's Use and Abuse of Power.

By Richard Thomas DeLamar. Dodd, Mead & Co., New York/Macmillan, London: 1986/1987. Pp.393. \$18.95, £14.95.

"How worried are you about the Japanese competition?", I asked a director of the British General Electric Company recently. "I'm not worried by Japan or any other country", he replied, "I'm only worried about IBM". It summed up the fear that many industrialists have about the computer giant — that it is now above the control of nations and can fortify itself to maintain a permanent semi-monopoly in the most critical industry of all.

How real is that bogey? Richard DeLamarter can present an authoritative and detailed view: for eight years he was a senior economist working for the Antitrust Division of the Department of Justice in Washington on their mammoth case against IBM — until President Reagan's new antitrust chief William Baxter abruptly dropped the case in 1982. DeLamarter was exasperated by the decision, convinced that IBM was exploiting its dominant position in computers against the national interest. "It is perilous", he says, "to trust a single company with such power". *Big Blue* is his own attempt to prove what the Antitrust Division were prevented from establishing.

It is a conscientious book, austerely and stiffly written, with little characterization or narrative drama and a good deal of repetition. But it provides unique evidence of how a giant multinational can exclude competitors in a fast-changing industry; and it bears the stamp of a professional investigator who has been exhaustive — even obsessive — in pursuit of his prey.

The most illuminating and vivid ingredients are the statements and memos, obtained through the antitrust case, from IBM's own executives and lawyers, as they strove to obliterate the competition while remaining legally correct. "Tell de Carlo" (an IBM memo instructed a marketing chief who was confronting rival manufacturers) "that we want him to blast anything and anybody out of his road who gets in the way of his accomplishing his mission." The IBM lawyers advised the use of much more circumspect language:

Avoid such phrases as 'containment of corporate threats' and substitute instead 'maintain position of leadership...'. Avoid references to the inexperience or naivete of IBM's customers or to their dependence on IBM.... Avoid implying that you might suppress technical

improvements unless competition forces your hand.

The lawyers' anxieties gave the clearest indication of what was going on. Behind the statesmanlike images of the great multinational — its dedication to progress, its concern to be a good citizen, its 'class act', as the author calls it — we have insights into the real and ruthless priority on which it had depended for its wealth: the destruction of competition wherever it emerged.

DeLamarter painstakingly records the means by which IBM has maintained its remarkable domination since its foundation in 1924. It was controlled by Thomas Watson Senior, who had already had his own searing experience of antitrust law.

"Behind the statesmanlike images of the great multinational — its dedication to progress, its concern to be a good citizen, its 'class act', as the author calls it — we have insights into the real and ruthless priority on which it had depended for its wealth: the destruction of competition wherever it emerged."

He had been a master-salesman for the National Cash Register company which had pioneered both cash registers and price discrimination to eliminate competitors; and in 1912 he was one of the first people to be convicted of restraint of trade under the new antitrust laws, after which he was fired from NCR. Watson was thus well suited to direct the new rival, IBM, which soon overtook NCR. By 1932 IBM had 85 per cent of the tabulating market, and it was sued by the antitrust division in the first weeks of Roosevelt's administration.

Computers provided far bigger scope, and a greater need to eliminate the competition. The author insists that IBM's own technology was frequently mediocre compared to that of its rivals, particularly when it hastily unveiled the 360 series mainframe computers in 1964. "Because we've been operating in a fortunate position all through our technological history compared with our competition", said Thomas Watson Junior, the son of the founder who also became the boss, "we have been able to generally take safe, conservative steps, and our percentage of wins has been exceedingly high". The 360 series was rushed to the market to forestall more enterprising innovators including Honeywell and Control Data. Once customers had installed the computers, they could be 'locked in' with software and interfaces which excluded competitors' products.

IBM steadfastly rejected the kind of standardization that would enable customers to connect up with other machines, and used its high profits periodically to

bring down prices in areas threatened by competitors. It might seem no more serious than Gillette's old trick of selling very cheap razors in order to commit shavers to their blades. But the IBM systems dominated a major industry, and the monopoly has (claims the author) "effectively frozen the commercial structure of the technically dynamic computer market".

This book provides a formidable indictment of IBM. Tantalizingly, however, it stops short in 1985, and most of the evidence comes from before 1982 when the antitrust case was dropped. To be fair, it was published last year in the United States; but the scant attention paid to the arrival of the personal computer in 1982 is disappointing, because the PC provides a critical new field of expansion in which to study IBM's behaviour. In some ways this story fits in with the author's thesis: the personal computer was first developed by the upstart Apple company, and IBM was compelled to set up an entirely new division to counterattack swiftly with its own PC. Apple then devised its new version, the Macintosh, and IBM have recently counterattacked again by announcing a new range with a new operating system which, in the usual IBM style, seems to be designed eventually to prevent rivals from being compatible with it.

It is doubtful whether IBM will be able to dominate the fast-growing area of personal computers as effectively as it has dominated the rest of the industry, and already its profits have taken a serious knock. But the real contest of the future, as DeLamarter suggests, is likely to be in the area of global communications systems, where IBM is competing with the giant telephone company AT&T as well as with other computer makers. Indeed, there are some signs that foreign competition, particularly from Japan, may be bringing an end to the fat years of IBM's monopoly; and the closer the computer becomes involved with global telecommunications, the more IBM will come up against constraints imposed by national governments. But this book reveals all the determination and ingenuity that IBM can mobilize in defending and extending its territory. Anyone who reads it will understand why those initials can conjure up such dread among the company's would-be competitors. □

Anthony Sampson, 27 Ladbroke Grove, London W11, UK, is a writer and journalist. Among other books he is author of *The Sovereign State: The Secret History of ITT and Black and Gold: Tycoons, Revolutionaries and Apartheid* (Hodder & Stoughton, 1973 and 1987).

New in paperback

• *The Background of Ecology: Concept and Theory* by Robert P. McIntosh. Publisher is Cambridge University Press, price is £12.50, \$16.95. The hardback edition was reviewed in *Nature* 318, 239 (1985).

Engines of change

John Forty

Supercomputers and Their Use. By Christopher Lazou. Clarendon: 1986. Pp.227. £22.50, \$45.

THE generally accepted definition of a supercomputer is a general purpose computer which is the most powerful computing engine available at any given time. In principle, therefore, there can only be one such machine but in practice a range of supercomputers is available and the choice depends on performance for the particular application. Inevitably the pace of advance in computer technology is such that no machine can hold the title for more than a few years; thus, in its day, the Ferranti Atlas was a supercomputer but was soon overtaken by the very powerful mainframes of the 1960s such as the IBM 360/195 and the CDC 7600.

In the past decade there have been enormous advances in computing power, defined loosely in terms of the ability to deal with the very large scale computation needed to solve complex problems, usually of a scientific nature. These have come about largely as a result of the development of new forms of computer architecture. Currently the most powerful

machines commercially available are the Cray and Control Data Corporation ranges, which exploit the concepts of multiple central processing and architectures which permit pipelining of a series of similar processing steps which can thereby be processed simultaneously, and consequently increase the speed of calculation. The latest Crays incorporate up to four powerful central processors so that this vector processing of data can be performed in parallel. Machines such as the ICL DAP (distributed array processor), consisting of a very large number of simple processing units working in parallel, are extremely effective in carrying out large scale computations where there is a large number of similar tasks to be performed simultaneously. Computing engines of this kind are orders of magnitude more powerful than the conventional mainframe.

A useful measure of progress over the past 20 years can be obtained by comparing peak performances, today's multi-processor machines, such as the 4-processor Cray X-MP48, being several thousand times faster than the Atlas. Advances in hardware technology have also resulted in a reduction of the cost per computational step by a similar factor. The result is an improvement in performance which is the equivalent of travelling at ten times the speed of Concorde compared with walking pace, for the same cost!

Computer design and engineering is continually evolving, and it might be thought that any account of supercomputers as we know them today could be of little permanent value. That may be so, but Christopher Lazou's little book does serve a useful purpose. It draws attention to the emerging scientific method of computation which is now seen by scientists to be a valid and technically feasible approach to solving problems (usually but not always of a scientific nature) which are so complex that the traditional approaches of experiment or analytical investigation are impractical. Supercomputers have made it possible to study the behaviour of the Earth's atmosphere and oceans in great detail, which will lead to a better understanding of climate, weather and so on. We can produce computer models of whole oil reservoirs to optimize output. And aircraft designers are able to design parts of, or even whole aeroplanes, and simulate the aerodynamic performance over a wide range of flight conditions with such realism that the time-consuming and expensive process of wind-tunnel modelling is no longer necessary. An important aim of the author has been to draw attention to the strategic importance of the supercomputer to the many areas of scientific research and engineering design in which it is now being applied.



Cray and his creation — Seymour Cray with the Cray-2 machine, announced in June 1985.

The basic concepts of supercomputers, the principles underlying their architecture and the developments in semiconductor technology which, of course, are fundamental to their performance, are set out in the first four chapters. The treatment is inevitably superficial, but the text and well-chosen illustrations make it easy to grasp the essential features. There are also some useful chapters on scientific computational methods, including scalar and vector techniques, and optimal vectorization of programs. The reader is introduced to the concept of performance in terms of hardware and software parameters such as the number of parallel processors, degree of vectorization and the fraction of work that can be processed in parallel. As its title declares, the book also covers the uses of supercomputers and a number of examples are described, drawn largely from the author's experience with the Cray 1 (the supercomputer of the late 1970s) at the University of London Computer Centre. Looking forward, there is a useful survey at the end of the book of new supercomputers known to be in the pipeline and of future developments that should be made possible by advances in technology.

Lazou writes in a fluent, informative manner. Although the subject is not explored in depth, the book provides a good introduction for those who are already computer literate and wish to know more about these developments at the forefront of computer technology. □

John Forty is Vice-Chancellor and Principal of the University of Stirling, Stirling FK9 4LA, UK. He is the author of Future Facilities for Advanced Research Computing, the report of a Joint Working Party of the Advisory Board for the Research Councils, the Computer Board and the University Grants Committee.

BOOKS ON NUCLEAR ENERGY

for the specialist
for the public

For our free catalog
call or write:

**AMERICAN
NUCLEAR
SOCIETY**

555 N. Kensington Avenue
La Grange Park, Illinois 60525
312-352-6611

Round and about the Universe

Michael Rowan-Robinson

Light From the Depths of Time. By Rudolf Kippenhahn. Springer-Verlag: 1987. Pp. 262. Pbk DM54.

I FIRST heard Rudolf Kippenhahn lecture at a summer school 20 years ago, and I remember the humour, clarity and imagination that he brought to his course on stellar structure. It is no surprise to me, therefore, that he should be capable of writing a really excellent popular book on astronomy. More surprising, perhaps, is the subject, cosmology. In fact the scope of *Light From the Depths of Time* is very wide — the structure and evolution of galaxies, the distance scale, special and general relativity, active galaxies and quasars, as well as the history of the Big Bang and the large-scale structure of the Universe.

The original, German edition of the book appeared in 1984 and the cosmological story presented dates from the early 1980s. Since then several developments have profoundly altered the cosmological consensus — grand unified theories, inflation, cold dark matter, biased galaxy formation — so this is not an account for those who want to read about the new synthesis of particle physics and cosmology. But this detracts little from the book's value as an introduction to cosmology and as an enjoyable read.

The qualities required of a good popular science book are imagination, a human dimension and a philosophical perspective. The main imaginative device employed by Kippenhahn is a series of dreams of one Herr Meyer, the man on the Munich omnibus. This device, frankly borrowed from George Gamow's *Mr Tompkins*, is very effective and takes the book to the brink of the truly inspiring (I do not fully understand why books by contemporary scientists are not able to cross that bridge of inspiration in the way that, say, James Jeans did, and that a writer such as Calvino does). Herr Meyer's dreams also add a human touch to the book, as, more conventionally, do the biographical anecdotes about astronomers.

The main virtues of *Light From the Depths of Time* are its vivid style and imaginative presentation, and these take it beyond the coffee-table type of book crammed with pictures which clogs up the astronomy shelf in bookshops (you know, that shelf underneath the eight shelves of astrology titles). It does not have a strong philosophical perspective, in the sense of a reason why the book had to be written. In his final pages, Kippenhahn inclines to the

view that the tenets of physics are projections of our own minds onto the Universe rather than inexorable laws that determine the course of events. I would like to have seen this argument developed further, perhaps at the expense of omitting the concluding device of compressing the history of the Universe into one year, with human history occupying the last fifteen seconds. This is simply Pascal's terror

before the infinite spaces and times of the Universe. But in a single human lifetime there is already more significance, more consciousness, more meaning than in all those vast, cold, empty spaces and times that exist without us. □

Michael Rowan-Robinson is Reader in Astronomy in the School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, UK.

Star attraction

Stephen P. Maran

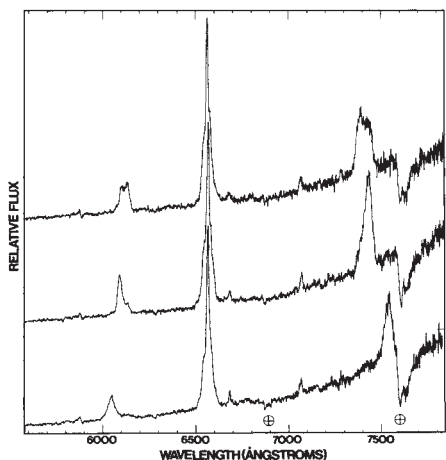
The Quest for SS433: The Discovery of the Astronomical Phenomenon of the Century. By David H. Clark. Adam Hilger: 1986. Pp.206. £9.95.

"Look at what we've been monitoring for Bruce" was an observer's excited request when I visited the control room of the 4-metre telescope at Kitt Peak National Observatory in October, 1978. "Bruce" was Bruce Margon, a young astronomer then at the University of California, Los Angeles, who had convinced several investigators to take spectra of the unusual star SS433 with telescopes in northern and southern California, and in Arizona, at every opportunity.

The spectra I saw on that occasion showed astonishing changes in the positions of emission lines from night to night. News of them was to burst upon the scientific world shortly thereafter, when Margon's group announced their discovery. SS433 is a bizarre star, which appears to have oppositely directed jets that spew matter into space at a quarter of the speed of light while they precess, tracing conical patterns in the sky once every 164 days. The jets arise from the compact member, perhaps a black hole, of an interacting binary system which has an orbital period of 13 days.

More than eight years have now passed, during which SS433 must have been inspected with every suitable optical and radio telescope in the world, and several satellites as well (Exosat and TENMA among them). Fresh details about its complex emissions are announced regularly, but we still have no confirmed reports of any truly comparable star, in our Galaxy or elsewhere. To the extent that there are known analogues for SS433, they are not stars but the active galactic nuclei that produce the powerful jets found in Cygnus A and other radio galaxies, and in quasars. SS433, on a stellar scale, is the closest example of the much more energetic phenomena in the hearts of those distant objects.

It was the British astronomer David H. Clark and his colleague Paul Murdin who, in June 1978, discovered the peculiar



Shifting spectrum — the moving emission lines (here at about 6050Å and 7400Å), the Doppler-shifted reflections of the hydrogen-alpha feature at 6563Å, which first indicated the bizarre properties of SS433.

spectrum of SS433, although not the changes in the emission lines. There were other astronomers and physicists who came close to making the key observations, but it was to Margon — who found, followed and identified the rapidly moving lines — that much of the glory went. Success came partly by luck, but largely from Margon's initiative and the ability to marshal the resources of several observatories equipped with large telescopes and blessed with good weather.

In *The Quest for SS433*, Clark tells the poignant and compelling tale of how close he came to revealing the secret of the star and how far others went in investigating it. It has been called a scientific detective story, but it also smacks of the thriller (although lacking a villain) as scientists make their way through international airports to observatories and meetings around the world, seeking clues to what *Sky and Telescope* later called the "enigma of the century".

Non-specialists can follow the story, but there are enough astrophysical data in the text and the illustrations to fascinate any scientist. Given the present job crisis in astronomy, this book should probably be kept out of the hands of bright young people — in case they should be unduly fired with enthusiasm for the subject. □

Stephen P. Maran is a senior staff scientist in the Laboratory for Astronomy and Solar Physics, NASA — Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

They came from space

John F. Kerridge

Cosmic Debris: Meteorites in History. By John G. Burke. *University of California Press*: 1986. Pp. 445. \$45, £35.95.

CONTEMPORARY books on meteoritics commonly devote little more than a few paragraphs to what went on before the twentieth century. A comment about the role of the Académie Royale des Sciences, an anecdote about Thomas Jefferson, a pat on the back for Chladni, and the reader is quickly plunged into the thicket of meteorite classification. Was there more to it than that? John Burke's meticulously researched book gives the resounding answer, yes. Here we may learn just what the Académie did or did not do, just what Jefferson did or did not say, and whether Chladni deserved his accolade — and, as the ads say, much more. The scholarship is first-rate and the style lucid. This is a book in which scientists and historians alike will find enlightenment and entertainment.

In discussing the progress towards recognition of the cosmic nature of meteorites, Burke argues that, contrary to the widely held view, the revolution was not accomplished overnight by the studies of Chladni, Howard and Biot but that "the chronicle of acceptance is...far more complex". Thus, he shows that Chladni's report on the Krasnojarsk pallasite, by eliminating local and anthropogenic origins through a series of plausible arguments, succeeded in convincing many that it, and other somewhat similar objects, had fallen to Earth as fireballs. However, despite the no less compelling logic of his contention that "it was just as reasonable to postulate small material bodies in space as to deny their presence", his proposal of a cosmic origin for such fireballs was not generally accepted until several decades later. One is reminded of the 'chronicle of acceptance' of plate tectonics, though Burke himself does not draw this parallel.

The struggle to impose a rational system of classification on the apparently chaotic and polyglot population of meteorites is described with clarity and scientific accuracy, as is the corresponding growth of sophistication in theories of their origin. Burke also brings out how improved resolution in classification proceeded at a pace set by developments in other branches of science, particularly chemistry and mineralogy. Indeed, an important message of this book is that from the eighteenth century onwards meteoritics has epitomized interdisciplinary science. Such interdependence transcends the purely

technical. Burke makes the point, for example, that acceptance of the cosmic origin of meteorites owed at least as much to the receptiveness of the prevailing cosmology as to the quality of meteoritic data. (In a similar vein, however, it is curious that the main developments in nineteenth-century geology apparently made no mark on the progress of meteoritics. Were meteorites too inherently anti-uniformitarian to fall within that emergent dogma?)

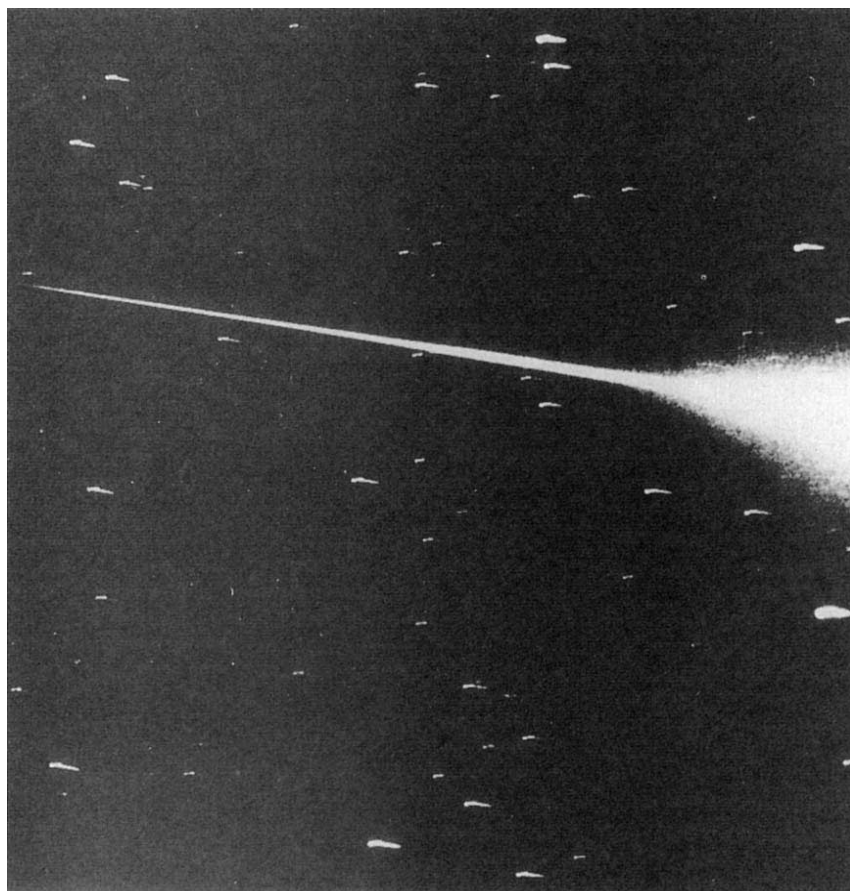
But if cosmology played a key role in fostering the scientific study of meteorites, information quickly began flowing in the opposite direction. Burke comments that, from 1860, "cosmologists could no longer disregard the evidence elicited from meteorites", though some meteoriticists doubt whether such a desirable state of affairs has been wholly achieved even by 1987. Nonetheless, the impact of meteoritics on cosmology has been considerable and nowhere is this clearer than in the emergence of certain primitive meteorites as our best exemplars of the cosmic abundances of the elements. Burke follows this exciting and fundamental line of inquiry as far as the work of Goldschmidt at which point, not unreasonably, he moves into geochemistry and models of the Earth. However, for the

continuation of the cosmic abundance story, Hans Suess's book, *Chemistry of the Solar System* (Wiley, 1987), is now available.

Burke concludes with a chapter on contemporary meteorite research and theories. In a book of this kind, such an account had to be selective. Not everyone will agree with all his choices of material, but that is inevitable and Burke nonetheless succeeds in capturing the dynamism and variety of current research. Given the historical perspective of the book, I would like to have seen some explicit discussion of the ideological context of recent studies of meteorites, in particular the dominance throughout the past 20 years of the nebular condensation paradigm. Also, in contrast to earlier chapters, this one contains some errors, albeit few in number and minor in importance.

In addition to the science, Burke supplies a wealth of cultural and sociological information to which I cannot do justice here, though it is a delightful aspect of the book. For example, how appealing it is to think that it took an H5 chondrite to end the Great Cultural Revolution! □

John F. Kerridge is at the Institute of Geophysics and Planetary Physics, University of California Los Angeles, Los Angeles, California 90024, USA.



Break-up over Bohemia — photograph taken in 1959 from a Czechoslovakian observatory at Ondrejov, near Prague, of the fragmentation of a bolide meteor. The body was estimated to be only 33,000 feet above the Earth's surface when it disintegrated.

The rich legacy of Erwin Schrödinger

P. W. Atkins

Schrödinger: Centenary Celebration of a Polymath. Edited by C. W. Kilmister. Cambridge University Press: 1987. Pp. 253. £30, \$54.50.

ERWIN Schrödinger was born a hundred years ago, on 12 August 1887. This volume contains the lectures delivered at the conference that was held in London at the end of March to celebrate his contributions to science.

Schrödinger's fame rests on his discovery of the equation that bears his name; this is a second-order partial differential equation, which depends on the masses of the particles and the potential they experience, and which underlies so many of the applications of quantum mechanics. Yet Schrödinger was active in fields much wider than the equation itself and his work spans almost the entire compass of fundamental science. We can trace the origins of gauge theory in his thoughts, and he made stimulating comments on the application of physics to biology.

Although his early life was spent in Austria, his first major appointment was in Zurich where he remained until he succeeded Planck in Berlin in 1927, shortly after the publication of his equation. In 1933, the year he shared the Nobel Prize with Dirac, Hitler's rise to power drove him, perhaps unnecessarily, from the country. He spent three unhappy, homesick years in Oxford before he returned to Graz in 1936. That too was unwise, for he arrived just in time to flee before the *Anschluss*, first to Rome and then, in 1940, to Eire, into the welcoming arms of de Valera. There he joined the newly created Dublin Institute for Advanced Studies, de Valera's brain child, and for 15 years was the director of its School of Theoretical Physics. In the end he returned to Austria and died in Alpach in 1961.

Schrödinger was 39 when he discovered his equation, an old fogey in comparison with most of his contemporaries (Heisenberg was 24, Dirac 24, Jordan 23, Pauli 25 and Einstein 26 when they were contributing to the subject most fruitfully) and almost too old for the task of adjusting to the transition from the old physics. Indeed, he clung to classical modes of thought, and originally proposed his equation as a classical theory of matter waves directly analogous to Maxwell's theory of electromagnetic waves. Questions of interpretation have been as much a part of Schrödinger's legacy as has been

the practical use of his formulation of quantum mechanics, and two of the contributions deal with aspects of this problem. Dorling explores what we would need to accept if Schrödinger's own interpretation of his equation were to be valid and examines the consequences of regarding it as a non-local theory. He sees the possibility that Schrödinger's interpretation may not only have been too hasty but that the potential of the Schrödinger



Schrödinger — more than the discoverer of the equation.

equation for calculating physical properties may have been underestimated.

Schrödinger was also shocked by quantum jumps, and hoped that his equation would evade them by replacing them with the smooth unfurling of the wavefunction. He was distressed by the realization that they remained intrinsic to his formulation ("If we have to go on with these damned quantum jumps, then I'm sorry that I ever got involved"). He would probably have been even more upset by Bell's tentative espousal of Ghirardi, Rimini and Weber's suggestion that, yes indeed, the wavefunction evolves smoothly according to the Schrödinger equation, but from time to time makes a jump, one that perhaps kills the cat. These cobbled theories are never very pretty when first proposed but, as Bell remarks, they may be a hint of better things to come. After all, since the Aspect experiment we are having to adjust ourselves to a more abstract,

non-local sense of the world.

Schrödinger's longest-lasting contribution to the mathematical description of the world may be less his equation itself than its implication that complex numbers are an intrinsic feature of our description of nature. Yang explores this point, which arises from Dirac's remark that "the real genius of Heisenberg and Schrödinger ... was to discover the existence of phase amplitudes". Yang points out that the importance of this discovery was not fully appreciated until the 1970s, with the development of gauge theories and the recognition of their relation to fibre bundles. There is now a general acceptance of the view that all fundamental forces are phase fields. Schrödinger's "almost casual" introduction of i into physics other than as a calculational device is the principal acorn planted by his work.

With the groundwork of interpretation examined, the contributions turn to a consideration of the features that can be calculated from the Schrödinger equation. Here I particularly enjoyed Thirring's survey of theorems relating to putative exact solutions of the many-electron equation for atoms and molecules, where we see how, contrary to their initial expectation, analysts seem to have progressed further with the many-body Schrödinger equation than with the corresponding classical problem. That is just as well, since the Schrödinger equation is the chemist's equivalent of the philosopher's stone: without the concepts it inspires and its calculational power chemistry would be hamstrung. A little of the efflorescence of this power is evident in Karplus's

account of the chemist's first, tentative calculation of the potential energy surface of the $H+H_2$ reaction trajectory burgeoning into their capacity to compute the dynamics of enzyme molecules containing thousands of atoms and surrounded by water molecules. In Buckingham's article we see how the ability of chemists to compute precise values using the Schrödinger equation has sometimes outstripped the experimentalist's precision, and has guided them to better interpretations of their data. Without the Schrödinger equation we would be unable to speak in terms of orbitals, and would be unable to face up to the sort of description of reactions that Fukui describes and which depend essentially on the language the equation provides.

However, these massive problems appear almost microcosmic when Hawking expands the concept to the Universe, and explores the possibility that when the

Mary Evans

Schrödinger equation is applied on this scale it eliminates the initial singularity in much the same way as its application to more conventional, smaller systems comes into its own at a point where classical physics breaks down. With this, the ultimate extension of the application of his equation to the Universe, Schrödinger would have been enthralled.

There is no space for me to examine all the papers in this volume, which includes contributions on boson condensation (Lewis), nonlinear optics (McConnell), unified field theories and particles (Hittmair, Salam), astronomy and astrophysics (Kibble, Seaton) and biology (Pauling, Perutz). In a sense that is best, for I must confess that I could not help being sympathetic to Pauling's somewhat egocentric contribution and in particular his view that Schrödinger's later work on biology, as recounted in *What is Life?*, was confused and misleading.

So much for Schrödinger the equation, what of Schrödinger the man? He remains elusive in these pages. Flamm explores his grandfather Boltzmann's considerable influence on Schrödinger. Despite the fact that they never met, Boltzmann was the spring of Schrödinger's interest in science: "His line of thought may be called my first love in science. No other has ever thus enraptured me or will ever do so again". Here are the roots of Schrödinger's attitudes and Boltzmann's work was often the point of departure for Schrödinger's own, including that on quantum mechanics. Their mutual admiration of Darwin was also the stimulus and support for Schrödinger's foray into biology, in which he sought, but more tentatively than Boltzmann, the physical basis of life. A clearer vision of the personality of de Valera, a thwarted academic who rechannelled his energies into politics, comes from McCrea's fascinating account of the emergence of the Dublin Institute and Schrödinger's involvement with it.

The scientific world directly and the general world indirectly has much to be grateful for in the life of Erwin Schrödinger. His was not so great a discovery that it would not have been made, perhaps within months, by others. Yet it fell to him to be first, and he had the power and confidence to explore many of its implications. That he left a rich legacy of calculational power and troublesome problems of the kind surveyed here is an epitaph he would have relished. These papers show that he reaches (as his later writings suggest he would like to think) from beyond the grave to tickle the intellect with the conundrums of quantum mechanics and to view the flourishing tree into which his acorn has grown. □

P. W. Atkins is University Lecturer in Physical Chemistry at the University of Oxford, and a Fellow of Lincoln College, Oxford OX1 3DR, UK.

Feasting with the Olympians

Walter Gratzer

The Life It Brings: One Physicist's Beginnings. By Jeremy Bernstein. *Ticknor & Fields, New York: 1987. Pp. 171. \$16.95.*

JEREMY BERNSTEIN is a theoretical physicist, turned writer. He must, one surmises, have been a pretty good physicist, for he broke bread with the likes of Yang, Lee and Gell-Mann; that he is a very good writer indeed is plain. His all-too-brief memoir began (like so much of the best science writing of recent years) as articles in the *New Yorker* and it is a wholly captivating read.

Bernstein conjures up for us his journey from a cultured middle-class Jewish home through one of the great New York high schools, cradle of Nobel Laureates (as for instance the St Elizabeth Gymnasium in Budapest once was in the intellectual heyday of Central Europe) and thence in 1947 to Harvard. There he slowly discovered his vocation by a process of trial and error that would be impossible in a British university and moved by way of pure mathematics into theoretical physics. He came under the influence first of Philipp Frank, student of Boltzmann and friend of Einstein, who embodied the European cultural tradition, and then of the remarkable indigenous school of younger physicists, led by the incomparable Julian Schwinger. It was Bernstein's first brush with genius. Schwinger was adulated and imitated by his students: to say "nuclear", it seems, marked you as one of his followers. His lectures (like those of Feynman on the Western seaboard) were the stuff of academic legend (which certainly percolated into other Harvard Departments, such as Chemistry, as I well recall): they were delivered without notes, and the subject was developed anew at the blackboard, impromptu, mathematics and all.

Schwinger was a prodigy, who had taught himself physics and mathematics from the *Encyclopaedia Britannica*. Bernstein relates the story of the adolescent's encounter with I. I. Rabi at Columbia. Rabi had summoned his then student, the astronomer Lloyd Motz, to his office to discuss a paper. Motz asked if he might bring in someone who was waiting for him outside. "He brought in this kid in knee pants", Rabi recalled. "So I told him to sit down some place, and he sat down. Motz and I were arguing, and this kid pipes up and settles the argument. And I said, 'Who the hell is this?'". Rabi eventually prised the lad away from City College, where he was failing courses in everything but physics, and with difficulty got him into Columbia.

Bernstein's later contact with the sinister world of the atom bomb and the dark contagion that seeped into physics from Los Alamos — the physicists, in Oppenheimer's words, had known evil — was prefigured by his acquaintance with Wendell Furry. Furry had been a communist, but heroically refused to "take the Fifth" when arraigned before the House Un-American Activities Committee, and was broken, like Oppenheimer, on McCarthy's wheel. His was, Bernstein thought, a maimed personality. There is a poignant vignette of Furry, who had in happier times been regarded as an outstanding theoretician, sitting in on Schwinger's lectures, turning over the pages of *Time* magazine.

From Harvard Bernstein progressed to the Institute of Advanced Studies, presided over by Robert Oppenheimer. It was the high noon of the Institute: T.D. Lee, C.N. Yang and a host of other luminaries were in residence, and Bohr, Dirac and Pauli were frequent visitors. "We will have a ball", Oppenheimer had said to Bernstein before he arrived. Bernstein attached himself to a genial young physicist, Marvin Goldberger (now President of Caltech), and together they tackled some problems concerning the interaction of fundamental particles. One of their col-

"Pauli was noted for intolerance and intellectual thuggery ('So young', he was supposed to have snarled as a student brought his research seminar to a close, 'and already he has accomplished so little')."

leagues was Freeman Dyson, of whose prodigious mathematical powers Bernstein speaks with awe. His first experience came when Dyson wandered into the room as Bernstein and Goldberger were brooding over an integral equation of seemingly unfathomable complexity. Invited to examine it, Dyson owned that he was feeling especially strong that morning and offered to have a go at it. Twenty minutes later he returned with the solution — later rediscovered and published by two other workers, whose name it bears. "I cannot imagine", Bernstein writes, "what it must feel like to think with that rapidity and clarity in mathematics. Does everyone else appear to be going in slow motion? But I have learned enough mathematics to get pleasure and delight each time I see it happening".

When the end of Bernstein's contract approached, Oppenheimer proposed a period in Europe with Wolfgang Pauli, but a lecture given by Pauli at Columbia persuaded Bernstein that the great man's powers were in a state of terminal decline (and indeed he died soon after at the age of 58). Pauli was noted for intolerance and intellectual thuggery ("So young", he was supposed to have snarled as a student



Jeremy Bernstein — pleasure in physics.

brought his research seminar to a close, "and already he has accomplished so little"). At the time he was preoccupied with a general theory of elementary particles, developed in association with Heisenberg, which was proving a delusion and which he was shortly to disown. He was confronted in the discussion that followed his talk by Niels Bohr, the most opaque and inarticulate of all the great *illuminati*. To capture the ascendant position — facing the audience — the two lumbering mastodons circled the lecture bench, locked into orbit. Bohr had made the characteristically luminous criticism that the theory, though crazy, was not crazy enough. It lacked the quality of what is these days called a paradigm-shift, that it should outrage the common sense of the day (or, as Oscar Wilde had it, man can believe the impossible, but man can never believe the improbable). Pauli demurred: his theory was quite sufficiently crazy. It was not, said Bohr. Freeman Dyson compared the scene to the death of some noble animal.

Bernstein went instead to Paris, impelled by nothing more than a *New Yorker* article on eating in France. He took to it like a duck to orange sauce and, with a physicist friend, ate his way through *Michelin*, learned French and met for the first time Murray Gell-Mann, who had attended the same high school but had so far outstripped his contemporaries as to graduate from university when still more or less a child. Gell-Mann surfaced in the laboratory as Bernstein and his French patron, Louis Michel, were formulating the problem that was meant to engage Bernstein's attention for the next year. Gell-Mann took it all in and left, only to reappear the next morning with the words — for he spoke flawless French — "Messieurs, le problème est résolu". He had developed a general framework that would accommodate the problem and

another of his own. Bernstein, awestruck, decided that an attempt must be made to understand the processes of thought that lay behind such casual miracles. He eventually conceded defeat, but he gives a transfixing description of the seemingly random explosions of perception that crackled around Gell-Mann's head. Genius, not for the first time, transcended analysis.

Bernstein expatiates on the pleasures of life in Paris in the early 1960s and on the anfractuosités of French science; he found modern theoretical physics flowering improbably among thickets of ancient and impenetrable prejudice. His reminiscences stop at the point at which he regretfully leaves Paris and begins to forsake physics for writing, for the rest of his experience, he felt, had already been processed through the writer's mill and needed no further journey of rediscovery. His account of a brief career in physics is rich in anecdote of the great and the eccentric, and it conveys with great immediacy the pleasures and the rigours of intense intellectual labour. Theoretical physics, perhaps uniquely among the natural sciences, has preserved the traditional intellectual values. Its practitioners are the Olympians of the scientific community. Their research cannot be divided

"In chemistry, in biology, and most of all in 'big' physics, success now too often accompanies skills of a kind more often associated with the less fastidious type of advertising executive."

up and delegated; collaborations spring up, because many theoreticians like to test evolving ideas by debate, but in Dirac's words, "the really good ideas are had by one person". In chemistry, in biology, and most of all in "big" physics, success now too often accompanies skills of a kind more often associated with the less fastidious type of advertising executive.

Bernstein's years in physics were enviably fulfilling. W.B. Yeats wrote:

*The intellect of man is forced to choose
Perfection of the life or of the work.*

Bernstein's book is something of a negation of this austere maxim. His title is taken from a letter that Robert Oppenheimer wrote to his brother Frank, in which he recommended a career in physics for "the obvious excellences of the life it brings". Bernstein's celebration of these excellences is not to be missed. □

Walter Gratzer is in the Medical Research Council Cell Biophysics Unit, King's College London (KQC), 26–29 Drury Lane, London WC2B 5RL, UK.

• On 22 May Princeton University Press will publish the first volume of *The Collected Papers of Albert Einstein*. The book, one of a projected total of about 40, is entitled *The Early Years*, and covers the period 1879–1902.



Gemstones

R.R. Harding & C.M. Woodward

Full colour illustrations taken from the premier collection at the Geological Museum, enhance this concise and authoritative guide to the subject.

64pp. Pbk. 0565 01011 5 £4.95

Classic Minerals of Cornwall & Devon

P.G. Embrey & R.F. Symes

Eighty specimens illustrated in full colour provide a spectacular record to many of the significant minerals to come from south-west England.

120pp, 80 colour plates, 30 figs.

Hbk. 0 565 00989 3 £9.95

The Alkaline Igneous Rocks & Carbonatites of the World Part 1 North & South America

A.R. Wooley

A series which will be of great interest to petrologists and geologists studying these economically important sources of many metals and non-metallic commodities.

224pp, A4.

Hbk. 0 565 00971 0 £40.00

Seaweeds of the British Isles

Vol. 3 Part 1

Fucophyceae (Phaeophyceae)

R.L. Fletcher

320pp, 16 plates.

Pbk. 0 565 00992 3 £30.00

Vol. 4

Tribophyceae (Xanthophyceae)

T. Christensen

48pp. Pbk. 0 565 00980 X £7.50

Further volumes in this comprehensive reference work on the marine algae of Great Britain and Ireland.

Encyclopaedia of Ferns Vol. 1

D.L. Jones

A comprehensive survey giving details of habitat, classification and cultivation. Will be of interest to both botanists and horticulturalists.

460pp, 250 colour photos.

192 b/w illustrations.

Hbk. 0 565 01019 0 £35.00

Sales Department, Natural History Museum Publications, Cromwell Road, London SW7 5BD

Physics before big physics

Laurie M. Brown

Rabi: Scientist and Citizen. By John S. Rigden. *Basic Books:1987. Pp.302. \$21.95.*
Alvarez: Adventures of a Physicist. By Luis W. Alvarez. *Basic Books:1987. Pp.292. \$19.95.*

ISIDOR Isaac Rabi and Luis W. Alvarez are both renowned experimental physicists, and are among the American science Nobel Laureates who rendered important public service during and after the Second World War. Rabi was largely responsible for the establishment in 1947 of Brookhaven National Laboratory on Long Island, New York — the first large national laboratory to be managed by a consortium of universities — and he has also been an international scientific statesman. At the General Assembly of UNESCO in Florence in 1950, he proposed, on behalf of the United States, a resolution calling for the establishment of a European physics laboratory, which was officially realized in 1952 as the European Laboratory for Nuclear Research (CERN) at Geneva. Alvarez's varied and adventurous career has included the development of nuclear weapons and of inventions important in military and civil aviation. An experienced pilot, he served as scientific observer on the B-29 bomber *Great Artiste* when it accompanied the *Enola Gay* on its fateful mission over Hiroshima on 6 August, 1945.

Rabi came from a penurious and unscientific, though culturally rich, background, while Alvarez's family was affluent and well-educated (his father, Walter, was a physiologist and a famous physician). Rabi was born in 1898, Alvarez in 1911, but Alvarez matured rapidly as a scientist and, at the outbreak of the Second World War, both men were in position to accept important tasks. Rabi became associate director of the Radiation Laboratory at the Massachusetts Institute of Technology under Lee DuBridge, working with British scientists, the military and industry on the development of radar devices. Later in the war, at Los Alamos, he counselled his friend Robert Oppenheimer, especially on personnel problems that baffled Oppenheimer. Alvarez worked at the Radiation Laboratory in a more direct way than Rabi, inventing radar systems for tracking submarines, for blind ground-control approach of landing aircraft (GCA), for early warning (MEW) and for blind-bombing (EAGLE). In 1943 he moved, first to Chicago to work with Fermi at the Metallurgical Laboratory, a cover name for the atomic bomb project, and then to

Los Alamos, where he applied his ingenuity to the implosion problem. Their experiences enabled both Alvarez and Rabi to assume positions of authority after the war, and they were among those who helped to transform physics research into big science — a milieu in which neither of them would be very satisfied to work nowadays.

The two books under review belong to a series, sponsored by the Alfred P. Sloan Foundation, which is intended to be essentially autobiographical. In that spirit, Rabi and his wife Helen collaborated closely with John S. Rigden, who is a well-known physicist at Washington University in St Louis, a scientific historian and the editor of the *American Journal of Physics*. Some of the material in his book was included in Jeremy Bernstein's article about Rabi that appeared in the *New Yorker* in 1970. But here Rigden provides welcome (though unobtrusive) historical documentation, so that we know not only what Rabi said, but when he said it, and to whom. Both Alvarez's and Rigden's writings are thoughtful and fresh. There is much insight in them, social as well as scientific, more explicitly in the Rabi book, more inferentially in Alvarez's. Both are accessible, and they deserve a wide audience.

Rabi's life is the stuff of which American urban legends are made. He was born Israel Isaac Rabi in Rymanow, in Galicia,

now Poland, a mostly Jewish town where Rabi, on a visit in 1968, could find no vestige of Jewishness — even the grave-stones had been levelled. (The name Israel later became 'Izzy', and turned into Isidor.) Rabi's father David, unskilled, and uneducated except in the Orthodox Jewish tradition, emigrated to the United States and soon sent for his wife Sheindel and their infant son. David Rabi worked at menial tasks, and the family lived in a Jewish ghetto, the fabled Lower East Side of Manhattan, full of saloons and synagogues, and thronged by prostitutes and criminals, as well as the pious. Home was two rooms, occupied by Isidor, his sister Gertrude, the parents and two boarders! There were visitors, whose discussions in Yiddish were often political, and generally socialist in tenor. Young Rabi developed rapidly in this environment — he could read the Yiddish newspaper before he started school at the age of five.

When Rabi was nine, the family moved to Brooklyn, where they lived in the back of a small grocery store. The whole family helped to run the store, and the money drawer was open to them all. Rabi says, "I didn't know we were poor". He discovered the local public library, and after reading all the children's books started on the adult section with a book on astronomy. Reading about the Copernican system was a turning point in his life. Rabi has always admired, and has had an emotional need for some coherent, all-encompassing intellectual system. Thus he moved from Orthodox Judaism to Marxist Socialism to Modern Science. His socialist views were abandoned, he says, when he looked around at his classmates in high school and thought, "These people could never run a country".

Rabi chose to attend Manual Training High School, walking three miles through gang-infested streets to get there, instead of a closer and higher quality school in a Jewish neighbourhood, in an attempt to escape from his Jewishness: "I had very definite notions of being an American in a broader sense". However, when he graduated in 1916 and went to Ithaca in upstate New York to attend Cornell University he "encountered a lot of discrimination and anti-Jewish sentiment". Likewise, when he graduated as a chemistry major three years later, neither he nor his few Jewish classmates were offered jobs in industry, although other chemists were "snapped up". There followed three years of unemployment, spent living at home and reading in the Public Library. Despairing of a job, he returned to Cornell as a graduate student in chemistry but discovered that the part of chemistry that he really enjoyed was called physics.

After a year, he left for Columbia University, where he invented an elegant null method for determining the magnetic susceptibilities of crystals. He also studied

Refraction of Beams of Molecules.

IN the Stern-Gerlach experiment the deviation of a beam of molecules in a magnetic (or electric) field is comparable to the optical case of the refraction suffered by a beam of light in traversing a medium, the refractive index of which varies in a direction perpendicular to the beam; the variation of the refractive index being analogous to the force or gradient of the field. However, in optical instruments the standard method of obtaining refraction is to allow the beam to travel from a homogeneous medium of refractive index n_1 to another of refractive index n_2 . The total refraction is then independent of the rate of variation of refractive index in the interface.

It is of interest to follow out the obvious analogy for a molecular beam. In the diagram, a beam of

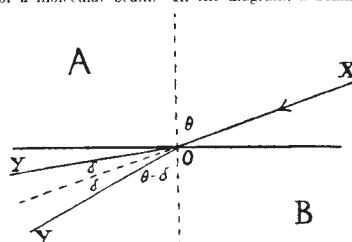


FIG. 1.

molecules XO passes from a region of no magnetic field A to another region B in which obtains a homogeneous magnetic field H perpendicular to the plane of the paper. Such a field can be produced between the flat pole pieces of a magnet. Let the beam be in the plane of symmetry between the pole pieces. We consider for simplicity a beam of alkali atoms in the normal state with kinetic energy E . The atoms will be orientated parallel or anti-parallel to H . Since there is no component of the force parallel to the edge of the pole piece, we have as in the optical case,

$$\sin(\theta - \delta) v_B = \sin \theta v_A$$

$$\frac{\sin \theta \cos \delta - \cos \theta \sin \delta}{\sin \theta} = \frac{v_B}{v_A} = \frac{\sqrt{2E}}{\sqrt{2E - \frac{m}{2} \frac{H^2}{4\pi^2 N_A}}}$$

Seminal paper — the opening half of Rabi's first report (Nature 123, 163; 1929) of work that was to win him a Nobel Prize.

the new quantum mechanics of 1925–1926, and with Ralph de Kronig solved the Schrödinger equation for the symmetrical top. After getting his PhD and marrying, he gave up the teaching job he had held at the City College of New York and went to Europe. Although he had not troubled to write ahead, in that golden age of physics he was able to visit Erwin Schrödinger at Zurich, Arnold Sommerfeld at Munich and Niels Bohr in Copenhagen. To Rabi's surprise, Bohr sent him, together with Yoshio Nishina from Japan, to Wolfgang Pauli in Hamburg. This was another turning point for Rabi, not because of Pauli, but because Otto Stern challenged Rabi to perform a molecular beam experiment that Rabi (then thinking of himself as a theorist) had proposed. In brief, the dénouement of this story is the following. In 1944, when the first Nobel Prizes in five years were announced, the physics prizes were awarded for molecular beam experiments; the 1943 prize went to Stern, and that for 1944 to Rabi "for his resonance method for recording the magnetic properties of atomic nuclei". Rabi's work forms the trunk of a whole tree of important scientific advances, many of them made by his students and co-workers, as Rigden points out.

Luis Alvarez never had to bother about social acceptance. His book is full of the pleasure that he takes in associating with influential people and in the perquisites of power and wealth. His friends are business and military figures, as well as the brilliant and honoured scientists he knows. He reserves his greatest admiration for activists, for inventors and organizers. He has flown a plane for over 1,000 hours, and he says that he thinks of himself "as having had two separate careers, one in science and one in aviation" and that he has found the two "almost equally rewarding".

Like other experimentalists of the school of his mentor, Ernest Lawrence, Alvarez feels strongly that theories of physics are not meant so much to be tested as challenged. When he did nuclear physics in the 1930s, he studied carefully a set of review articles that made up what came to be called (Hans) "Bethe's Bible", about which he remarked later, "If [Bethe] said a phenomenon would never be observable, I wanted to prove him wrong". Among his pre-war achievements, he demonstrated electron capture by atomic nuclei and the decay of H^3 into He^3 . (It had been thought that H^3 was more stable than He^3 .) In 1968 he received the Nobel Prize for developing the hydrogen bubble chamber and associated analytical techniques that allowed the discovery of many resonance states of strange particles.

In the book, Alvarez offers his views on the present status of the fields of physics in which he has been most prominent.



End of a phase — Luis Alvarez on his last day at the MIT Radiation Laboratory, September 1943.

"There is no way", he says, "that a person with my personal qualities could go either into nuclear physics or particle physics at the present time". One of his criticisms is that experiments are carried out by such large groups, reaching to hundreds of people, that he would miss the pleasures of essential participation and of individual recognition. His second criticism is what he considers to be an undue influence of theorists in making funding and scheduling decisions at particle accelerators. About experiments that merely check what "every theorist was convinced" of, such as the discovery of the W and Z particles at CERN, he asks, "What was really learned?"

Alvarez and Rabi both witnessed the first nuclear bomb test at Alamogordo, and found it to be a profound experience. Both are convinced (as one might expect of good establishment scientists) that it was morally justified not only to make the bombs but to drop them on Japan because, they argue, on balance Japanese and Allied lives were saved by those acts. (I have never been comfortable with this quantitative approach to mortality or morality.) Alvarez and Rabi differed, however, after the Soviet Union's first nuclear explosion on 29 August, 1949, Alvarez joining Lawrence and Edward Teller in proposing a crash programme to build the "Super" (as the earlier proposed version of the hydrogen bomb was called), while Rabi concurred with the rest of the General Advisory Committee to the Atomic Energy Committee, headed by Oppenheimer, that such an action was unjustified at the time. It would be unfair to try to summarize their reasons here. At the hearing which led to the suspension of Oppenheimer's security clearance in 1954, Rabi appeared in his defence, while Alvarez testified that he thought Oppenheimer's reasoning on the Super had been faulty, but did not question his

loyalty to the United States.

Rabi's main contributions, other than his science, may turn out to have been in influencing Eisenhower to establish a strong science advisory presence in the White House and his efforts on behalf of international control of atomic energy (which led to the First International Conference on the Peaceful Uses of Atomic Energy in August 1955). Alvarez himself feels that he may be best remembered for having proposed, with his geologist son Walter, the theory that the dinosaurs and some other species of animal became extinct as the result of the collision of the Earth with an asteroid about ten kilometres in diameter, causing huge clouds of dust to obscure the Sun. This has led some physicists to consider the possibility that a nuclear war might give rise to similarly large dust clouds and result in the so-called "Nuclear Winter". If this should lead to rethinking on the part of those who still consider nuclear war as a survivable alternative for mankind, then Luis and Walter Alvarez may have, indirectly, made a notable contribution to world peace. □

Laurie M. Brown is Professor of Physics and Astronomy at Northwestern University, Evanston, Illinois 60201, USA. Twenty-five years after I.I. Rabi, he attended Manual Training High School in Brooklyn and Cornell University.

The Quantum and Beyond

William M. Honig

Western Australian Institute of Technology, Perth, 6102, Australia.

MAX JAMMER 'Thought-provoking important new ideas on fundamental issues, intellectually courageous'

FRANCO SELLERI 'Inspiring book ...for the deeper understanding of subatomic reality on a causal basis'

EMILIO PANARELLA 'Wealth of new concepts, aesthetic beauty, powerful new analytic tools...an intellectual treat'

MAURICE SURDIN 'A courageous effort to go beyond the established'

LUDWIK KOSTRO 'Hydrodynamical models offer many possibilities'

SIMON PROKHOVNIK 'Useful and intelligible view of $E = hf$ via photex model'

\$17.50 Soft Cover, \$25 Hard Cover
317 pages, 44 illus, Postage \$3.50.

Philosophical Library, Inc.
200 W.57 St., NY, 10019, NY

Rain, rain, rain

Malcolm Walker

Monsoons. Edited by Jay S. Fein and Pamela L. Stephens. Wiley: 1987. Pp. 632. £71.75, \$104.95.

DERIVED from the Arabic word *mawsim*, meaning fixed season, the term monsoon refers strictly to the seasonal nature of the winds which blow over southern Asia, Indonesia, northern Australia, the Arabian Sea, the Bay of Bengal, the South China Sea and much of North Africa. Popularly, however, the word refers also to the summer rains of these regions.

In all monsoonal regions, abrupt changes in atmospheric behaviour take place at specific times of year. Characteristic of the onset and retreat phases of monsoons, these changes involve comprehensive and widespread reorganizations of atmospheric circulation patterns. Reversals of wind direction occur at all levels in the troposphere and lower stratosphere; and associated changes in patterns of vertical motion are such that the droughts of winter are replaced by wet weather in summer. Rainfall varies markedly from year to year. In some years it is widespread and abundant. In others it is scanty. It is always unreliable. Sometimes the rains fail at a critical time; and sometimes there are short bursts of torrential rain rather than the periods of steady rain preferred by farmers.

Monsoons are of enormous social, agricultural and economic importance, and they dominate the lives of more than a thousand million people. For these people there are vital questions. Will the monsoon be early or will it be late? Will there be too much rain or too little? Meteorologists are interested in these questions too, and they seek answers to related questions, partly for altruistic reasons and partly in response to a purely scientific challenge. What drives monsoons? Why do they vary so much from year to year? Can they be forecast? How do they interact with atmospheric circulations outside the tropics?

Natural philosophers have long been fascinated and occupied with monsoons, particularly that of southern Asia, the most impressive of all seasonal phenomena. Indeed, the classical thermal concept, which likens a monsoon to a giant sea breeze, and which is still widely accepted even though its shortcomings have been recognized for many years, was first proposed as long ago as 1686, by Edmund Halley, in a paper published in *The Philosophical Transactions* of the Royal Society of London.

Until about 30 years ago reliable investigations of monsoons were impossible because of the paucity of upper-air

data and the consequent need to base the study of tropical weather systems largely upon surface data. Since then, however, techniques have been developed which enable us to reach well into the stratosphere. Moreover, satellites have provided a new source of data, and computer power to process the data has increased greatly. There has also been a growth of interest in the tropical atmosphere, from which monsoon research has benefited; and a number of special research enterprises involving international collaboration have been mounted in monsoonal regions (the International Indian Ocean Expedition of 1959–1965, the Indo-Soviet Monsoon Experiment of 1973 and the International Monsoon Experiment of 1977–1979, for example). Advances in monsoon meteorology have been rapid in recent years, so it was appropriate to pause and take stock of current knowledge.

This book is more than a scientific review, however, for it is concerned not only with atmospheric processes and the wind-driven circulations of the Indian Ocean but also with various other facets of monsoons. There are, for example, three chapters devoted to the literature and folklore of India; these accounts deal with poetry and prose, proverbs, symbolism, religion, rituals, festivals and medical regimens, and they demonstrate the enormous and continuing influence of the south-west monsoon on virtually every aspect of traditional culture in India. There is, too, a chapter on ancient and mediaeval records of the monsoon winds and currents of the Indian Ocean; and in another contribution the evolution of monsoon meteorology is traced from the seventeenth century to the twentieth. Other chapters concentrate upon the economic implications of weather and climate in monsoonal regions, the economic consequences of abnormal monsoons, and the strategies and programmes by means of which governments might deal with the aftermath of such monsoons.

Ten of the 19 chapters are devoted to meteorology, with the emphasis upon the vagaries and variability of monsoons on various time scales. One chapter is concerned with interannual and intraseasonal variability, for example, and another considers the evidence for and possible causes of secular changes in the timing, extent and strength of monsoons. The problem of prediction also receives a great deal of attention. The characteristics, applicability and limitations of various types of numerical models are discussed at length; and two chapters focus upon the mix of synoptic, empirical and numerical techniques used operationally by weather forecasters in India and South East Asia.

It is clear from the book that monsoons are meteorologically much more complex than was supposed only three decades



After the deluge — monsoon flooding of field systems on the plains of India.

ago, let alone in Halley's time. It is also clear that a complete understanding of them cannot be achieved without a thorough grasp not only of the underlying causes of atmospheric circulations and the relevant physics, but also of orographic influences, the effects of sea-surface temperature anomalies, and the teleconnections between monsoons and atmospheric events far from monsoonal regions. Indeed, monsoons are evidently not so much regional phenomena as elements in the global ocean-atmosphere system. Numerical modelling of tropical weather systems has plainly not yet advanced much beyond the experimental stage, and further progress in monsoon forecasting on all time scales will depend upon improved models and even greater computer power. More data will also be needed, as observations, unlike ideas and models, are still not plentiful.

The book is intended for meteorologists, climatologists, oceanographers, environmental scientists, folklorists, administrators and policymakers, and it takes the form of a collection of essays, each written by a distinguished expert. The 16 authors — and the editors too — have acquitted themselves well. The style is lucid and fluent throughout; every chapter includes an extensive bibliography; numerous illustrations and tables of data are provided; and the standard of presentation is consistently high.

Although the approach is largely non-mathematical, a familiarity with the physics of fluids is required to follow some of the meteorology. Nevertheless, the general reader should not be deterred. As claimed in the blurb, this volume will be an "an ideal sourcebook for anyone interested in a phenomenon that so greatly affects so much of mankind". □

Malcolm Walker is a Senior Lecturer in the Department of Maritime Studies, UWIST, Cardiff CF1 3EU, UK.